IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

ScarNet: Development and Validation of a Novel Deep CNN Model for Acne Scar Classification with a New Dataset

MASUM SHAH JUNAYED[1,2](MEMBER, IEEE), MD BAHARUL ISLAM[2,3] (SENIOR MEMBER, IEEE), AFSANA AHSAN JENY[2], AREZOO SADEGHZADEH[2], TOPU BISWAS[4](MEMBER, IEEE), A. F. M. SHAHEN SHAH[5] (SENIOR MEMBER, IEEE)

[1]Department of Computer Science and Engineering, Daffodil International University, Dhaka-1207, Bangladesh
[2]Department of Computer Engineering, Bahcesehir University, Istanbul 34349, Turkey
[3]College of Data Science & Engineering, American University of Malta, Bormla 1013, Malta
[4]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia
[5]Department of Electronics and Communication Engineering, Yildiz Technical University, Istanbul 34220, Turkey

Corresponding author: Masum Shah Junayed (e-mail: masumshahjunayed@gmail.com)

This work is partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under 2232 Outstanding Researchers program, Project No. 118C301.

ABSTRACT

Acne scarring occurs in 95% of people with acne vulgaris due to collagen loss or gains when the body is healing the damages of the skin caused by acne inflammation. Accurate classification of acne scars is a vital factor in providing a timely, effective treatment protocol. Dermatologists mainly recognize the type of acne scars manually based on visual inspections, which are time- and energy-consuming and subject to intra- and inter-reader variability. In this paper, a novel automated acne scar classification system is proposed based on a deep Convolutional Neural Network (CNN) model. First, a dataset of 250 images from five different classes is collected and labeled by a well-experienced dermatologist. The pre-processed input images are fed into our proposed model, namely *ScarNet*, for deep feature map extraction. The optimizer, loss function, activation functions, filter and kernel sizes, regularization methods, and the batch size of the proposed architecture are tuned so that the classification performance is maximized while minimizing the computational cost. Experimental results demonstrate the feasibility of the proposed method with accuracy, specificity, and kappa score of 92.53%, 95.38%, and 76.7%, respectively.

INDEX TERMS Acne scars, dataset, image classification, CNN, skin disorder, skin image analysis

I. INTRODUCTION

Acne scars originate in the tissue damage resulting from the inflammatory acne lesions and Vulgaris, which may need long-time treatment to be removed [1]. Consequently, acne scars can be more troublesome for patients than acne vulgaris, and they may have substantial adverse psychological and emotional effects on the social life of patients. It weakens the quality of life and can lead to anger, depression, embarrassment, poor self-image, low academic performance, social withdrawal, and limited employment chances [2]. It is reported in Global Burden of Disease Project [3] that the prevalence rate of acne is 9.4% of the global population and identified as the 8^{th} most common disease the worldwide [4]. Almost over 90% of adolescents (i.e., 95-100% of boys aged 16-17 and 83-85% of girls aged 16-17) experience acne based on hormonal fluctuations, and it persists into adulthood (i.e., adults in their 40-the 50s) in about 12-14% of cases [5]. Due to the lack of dermatologists and the long-awaited time for an appointment (an average of 32 days) [6], patients do not receive timely acne treatments. This delay in starting the treatment process is permanent, disfiguring acne scars because of some injuries in the sebaceous follicle.

There are numerous options available for acne scar treatment, such as chemical peels, laser treatment, punch techniques, cosmetic filler, dermabrasion/microdermabrasion, microneedling, onion extract, local tumescent, subcision, surgery, and combined therapies [1]. However, choosing the effective and appropriate option and its related protocols for ideal acne scar treatment is vital first to recognize the acne scar types and subtypes. Although several works have been

reported for acne scar treatment in the literature, there have been only a few descriptive terms to diagnose the types of acne scars [1], [7], [8].

There are three main types of acne scars: atrophic, hypertrophic, and keloidal, among which atrophic is the most well-known type [7]. Atrophic scars are further subdivided into icepick, rolling scars, and boxcar [9], which results in totally five acne scar classes. For clinical purposes, dermatologists usually classify the acne scars into these five types based on a simple visual inspection (using the descriptive terms) by a direction or overhead light, a magnification/dermatoscope, and identification marks on patients' faces. Two quantitative scoring methods have also been proposed for acne scar grading [10], [11]. However, in these visual inspection-based methods, the classification process is time-consuming, has high error rates, and causes discomfort for patients. Additionally, as the descriptive methods are based on the textural irregularities and the physical characteristics of the scars, such as colors, shapes, patterns, width, depth, 3D architecture, etc., the classification results are highly subjective to the expert's experience and knowledge. Hence, to infer consistent treatments for acne scars, a unique accepted standard acne scar classification system is required to recognize the scar type efficiently and fast, with low cost and minimal dependence on expert help.

In the last decade, deep neural networks and machine learning techniques for medical image analysis, especially skin disease detection and classification such as acne vulgaris, cancer, etc., have received much attention [12]–[21]. Even a few studies were conducted for acne scar detection [22], [23]. However, there have been no AI-based automated acne scar classification systems to the authors' best knowledge. Generally, providing such a system is challenging due to several main reasons: (a) existence of many different human skin tones, (b) large variations in the appearances of acne scars such as their shape, size, and even the position of the scar, (c) dependency of the acne scars on the patients' age, gender, and skin types, and (d) the lack of an appropriate and publicly available dataset with a large number of images sufficient for learning-based methods (previously published datasets are only for acne vulgaris classification and grading [24]–[29], not for acne scar).

An acne scar dataset (with five classes) and an automated acne scar classification system based on a deep convolutional neural network (CNN) are proposed to address the issues mentioned above and the challenges. Overall, the contributions and strengths of this work can be summarized as follows:

- An acne scar dataset named *"5-class Acne Scar"* has been created, including 250 images labeled in five classes of acne scar types by an experienced ophthalmologist. To the authors' best knowledge, *"5-class Acne Scar"* is the first dataset in this field that provides the images from five acne scar types of Hypertrophic, Keloidal, Icepick, Rolling, and Boxcar.
- As the first attempt, an automated computer-aided acne scar classification system is proposed based on a novel deep 19-layer CNN model, i.e., *ScarNet*. In this model, the activation function, optimization algorithm, loss function, kernel sizes, and batch size are adjusted to reduce the computational cost, such as model size, training parameters, and running time. At the same time, the accuracy is still high with a limited number of layers compared to the pre-trained models.
- To obtain the finest mass images, our dataset passes a pre-processing step, including image resizing, quality improvement, and augmentation. We benefit from the guided filter as a contrast enhancement technique to reduce the noise while preserving the edges of the acne scars. Additionally, to have a better representation of the acne images and approximate the human vision, $L^*a^*b^*$ color space is applied, which results in competitive classification performance. Pre-processing step ends with augmentation, which increases the number of training samples five times to prevent model overfitting.
- According to the extensive availability of smartphones, the proposed system can be used by ordinary people, especially in underdeveloped countries (where there is a lack of dermatologists) as a remote screening system.
- Comparing the performance of *ScarNet* with four conventional machine learning (ML)-based classifiers, i.e., Decision Tree (DT), Multi-layer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest, and five pre-trained deep learning-based models, i.e., Inception-V3, MobileNet, RedNet-50, AlexNet, and VGG-16 on the developed dataset, our model attained competitive performance in acne scar classification with high accuracy and minimized computational cost.

The rest of the paper is organized as follows. In Section II, the recently reported studies related to skin diseases, especially acne vulgaris detection and classification, are reviewed. Data collection is presented in Section III. Section IV represents our proposed system in detail, including the pre-processing step and the architecture of our CNN model. The evaluation metrics, experimental setup and analysis, comparisons, and the related discussions are all presented in section V. Finally, section VI concludes the paper and provides our future research directions.

II. RELATED WORKS

Although numerous approaches have been reported in the literature for skin image analysis, such as detection and grading of acne vulgaris, skin cancer, skin eczema, etc., in the last decade, the literature lacks computerized systems for automated acne scar classification. Generally, skin disease-related studies can be categorized into two main groups: classical computer vision- and machine learning-based methods and deep learning-based approaches. Some of the acne-related approaches are discussed briefly in this section.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

A. CONVENTIONAL APPROACHES

In conventional methods, the handcrafted features are extracted from the raw images and then employed for different processes such as detection and classification. In [30], acne and facial skin defect recognition system was proposed based on SVM classifier. After extracting the region of interest (ROI), particular color space and an adapted marginal area were used to distinguish possible defects from the ROI. Finally, specific important texture characteristics from 126 acne, 134 marks, and 134 regular skin patches were employed in the final classification of the experiments. Although it received an accuracy of 99.40%, it is time-consuming due to including multiple stages.

Dey et al. [23] proposed a system for detecting acne scar pixels based on color image processing. In this method, the knowledge base was created by recording pixels from skin as background and from acne scars as ROI. The segmentation was carried out using Mahalanobis distance (MD) based on the minimum-distance rule in the RGB space. Applying MD, average values of 90.36 and 93.82 were obtained for sensitivity and specificity, respectively. An acne grading method was presented in [31] based on SVM, which was applied to classify acne vulgaris into four classes, i.e., mild, moderate, severe, and very severe. Khongsuwan et al. [32] proposed a counting system for estimating the number of acne vulgaris as points on the skin whose prediction accuracy was 83.75%. Later, a mobile application was developed in [24] for acne vulgaris detection, classification, and segmentation evaluated on only 35 images in five classes. A binary classification (normal skin and acne) was performed based on Fuzzy C-Means (FCM) accuracy of 100%. Another binary classification task was conducted for distinguishing acne scar from inflammatory acne based on SVM and FCM with an average accuracy of 80% and 66.6%, respectively. However, for a proper evaluation, more images are required. Abas et al. [33] proposed a method for counting the number of lesions on a region of skin and segmenting them into six classes (5 classes of acne vulgaris types and one scar class). In their method, ROI is extracted based on entropy-based filtering and thresholding. Applying a binary classification tree and a combination of two descriptors, i.e., discrete wavelet and gray-level co-occurrence matrix, an accuracy of 85.5% was achieved.

An acne detection approach was presented in [34] based on applied speeded up robust features with an average accuracy of 68%. Recently, a hybrid acne vulgaris classification method has been proposed by Hameed et al. [27] based on Naive Bayes Classifier (NBC) and image processing on a dataset of 3 classes with 40 images in each. Their system obtained an accuracy of 93.42%. Generally, using machine learning techniques such as K-means clustering and SVM is complicated and challenging as the handcrafted features must be extracted before training the model. Additionally, the feature extraction and arrangements are task-subordinate and cannot be summed up easily. These methods also suffer from the lack of self-adaptivity due to manual feature extraction. Although many conventional approaches have been published for acne vulgaris detection and classification, as far as we know, there are still no studies for acne scar classification, even based on these classical techniques.

B. DEEP LEARNING-BASED APPROACHES

According to recent studies, the extracted features based on CNNs are deeper and more effective than hand-crafted features extracted by conventional techniques. Inspiring from this fact, the outstanding performance of CNNs, significant advancement in GPUs, and the availability of large datasets, deep neural networks have gained growing popularity and attracted much attention among researchers in medical image analysis. In [25], the skin images were first distinguished from non-skin images by applying a binary-classifier. Then seven different types of acne vulgaris were classified based on the features extracted from a CNN-based model. AcneNet was proposed by Junayed et al. [26] based on a Deep Residual Neural Network to recognize five types of acne vulgaris using a total number of 1800 images. Compared to two pre-trained models, i.e., Inception-V3 and MobileNet, their model received higher accuracy of 86.28% and 95.89% for training and testing, respectively, which shows a noticeable amount of underfitting in the training process.

A facial acne Vulgaris grading system was proposed by Zhao et al. [35] using 4,700 selfie images labeled in 5 classes of severity. First, the irrelevant background was removed to find facial landmarks through which the key skin patches were selected. Using the pre-trained ResNet-152 and transfer learning approach, the features were extracted from the images, fed into a fully connected layer for final classification. Chitransh et al. [36] utilized five classifiers such as Logistic Regression, Random Forest, Kernel SVM, Naive Bayes, and CNN to detect three different types of skin diseases, precisely, Acne, Lichen Planus, and stevens-johnson syndrome (SJS) and toxic epidermal necrolysis (TEN). Among these classifiers, CNN performed better than other machine learning algorithms. The accuracy of Logistic Regression, Random Forest, Kernel SVM, Naive Bayes, and CNN were 73.76%, 73.36%, 50.7%, 49%, and 99.05% respectively.

Junayed et al. [37] proposed a deep CNN model to recognize and classify five types of Eczema diseases using 2500 images with an accuracy of 96.2%. Despite the high accuracy, it suffered from overfitting as around 24M parameters were utilized in their proposed model with no dropout layers. The acne vulgaris classification performance of the Faster Region-based CNN (Faster R-CNN) and Region-based Fully Convolutional Network (R-FCN) were compared by Rashataprucksa et al. in [29] using 871 images in four classes. A mean average precision of 28.3% was obtained for R-FCN, which is still low for real-life clinical applications. Although the deep learning-based methods have been extensively developed and applied in medical image analysis (i.e., skin disease detection and acne vulgaris detection, grading, and classification), there is still a lack of an automated acne scar classification system.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

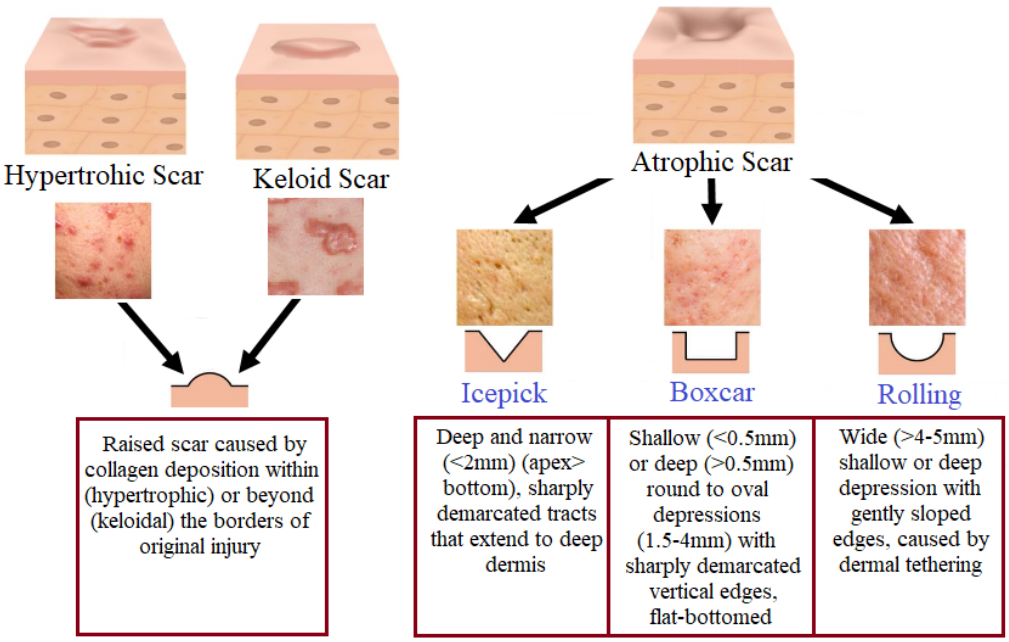

FIGURE1: Samples from five different types of acne scars and their characteristics, i.e. Hypertrophic, Keloid, Icepick, Boxcar, and Rolling scars.

III. DATASET DEVELOPMENT

Acne scars are the secondary lesions of acne vulgaris formed by the natural reaction of body tissue while healing the existed injuries. Depending on the net loss or gain of collagen in the tissue in the healing process, there are two basic types of acne scars [7]: atrophic and hypertrophic/keloidal scars, respectively. The most common are atrophic acne scars, further divided into three sub-classes of rolling, icepick, and boxcar. All these five types are illustrated in Fig. 1 and their main characteristics are presented as follows:

Rolling: they are typically more extensive than 4 to 5 mm. Having a varying depth, these spots create an "M" shape with sloping edges giving the skin a wavy and uneven appearance.

Icepick: they are "V"-shaped, more narrow, and profound, pointing down into the skin, which mostly appears on the cheeks. The treatment of this type of scar is callous and requires a continuous process.

Boxcar: they are broad, box-shaped scars like a "U" with a wide base and sharp vertical edges mostly caused by chickenpox, virus-based rash, widespread acne vulgaris, or varicella. They commonly appear on the thick part of the skin, such as the lower cheeks and jaw.

Hypertrophic: typically, these scars are pink, raised tissue on the skin that stays inside the actual acne's boundaries, which means they have the same size as the acne from which they are originated. They commonly appear on the darker skin color on the chest, shoulders, back, and jawline areas.

Keloidal: these are raised scars very similar to hypertrophic. The only difference is that keloidal spots are more significant than the initial acne and propagate beyond their limits.

The degrees of all types of acne scars are increased if they receive delayed treatment [38]. There have been numerous treatment methods for acne scarring, among which the most effective and suitable treatment options can be selected based on the type of the scar [8]. For instance, Rolling acne scars can be treated by the picosure laser or trichloride acid peel, while punch excisions and AFL are helpful and effective treatments for Icepick scars [39]. It is also true about home treatment. For example, atrophic acne scars can be successfully treated at home with topical over-the-counter (OTC) retinoids, such as Differin. In contrast, hypertrophic and keloid scars are softened, and their height is reduced using gel silicone sheets. Hence, it is vital to have an automated acne classification system with highly accurate performance for timely treatment.

One of the significant and challenging factors in applying computer-aided techniques, especially the deep learning-based ones, to clinical applications is getting access to a proper dataset with sufficient images already labeled by a health professional. There is still no publicly available dataset for acne scar classification in the literature to the authors' best knowledge. To overcome this challenge and fill this gap between the research and medical applications, a dataset of 250 RGB images in 5 different classes of acne scars (each class contains 50 images), namely *"5-class Acne Scar"*, is proposed. Among them, 51 images have been captured by the authors from the visited patients to the Dhaka Medical College (DMC) hospital by a 48-MP camera of a smartphone. These data were collected by following the individual patient's consent. The remaining 199 images were gathered from DermNet [40]. A well-experienced dermatologist has

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

FIGURE2: Sample images from our developed acne scar dataset *"5-class Acne Scar"*, the rows from top to bottom represent the acne scar classes of rolling, icepick, boxcar, hypertrophic, and keloidal, respectively.

labeled all images in five different classes. Some samples of these five classes in our dataset are illustrated in Fig. 2.

IV. PROPOSED SCARNET METHOD

The main flowchart of the proposed acne scar classification system is illustrated in Fig. 3. First, the images of our developed dataset are passed a pre-processing step. The pre-processed images are fed into a CNN-based model for feature extraction as well as classification. The details of each step are presented in the following subsections.

A. PRE-PROCESSING

The quality of the input images can significantly affect the performance of the whole system. Hence, image pre-processing is essential to improve the image quality and have more accurate subsequent image analysis processes such as classification. The pre-processing module comprises three sub-modules in our proposed method: image resizing, image quality improvement, and augmentation. Before applying the two latter sub-modules, the images of the developed dataset are all resized to 224×224 pixels.

Image quality improvement. One of the main factors in medical image analysis is improving the contrast of the images for providing a better representation. Hence, in our proposed system, guided image filtering [41] is employed as contrast enhancement to reduce the noise of the input images while preserving the edges of the acne scars. This enhancement in the edges of the scars makes it easier for the model to train different acne scars and distinguish between them. The guided filter is similar to conventional bilateral filter [42] as an edge-preserving smoothing operator but with a difference that it has a better performance close to the edges with a fast and non-approximate linear-time algorithm. In this algorithm, the filtering size has no impact on the computational complexity, and the performance is outstanding both in quality and efficiency. It is assumed that there is a linear relationship between the guidance I and the filter output q since the object edges depend on its gradient. If q is a linear transformation of I in a window w_k (k is the center pixel), the basic model of the guided filter is as follows [41]:

$$\begin{aligned} q_i &= a_k I_i + b_k, \forall i \in w_k \\ q_i &= p_i - n_i \end{aligned} \tag{1}$$

where (a_i, b_i) are some constant linear coefficients and w_k, q_i, p_i, n_i, and I_i are output pixel, input pixel, pixel of noise components, and guidance image pixel, respectively. As $\bigtriangledown q = a \bigtriangledown I$, it is guaranteed that q has an edge only if I contains an edge. Some samples of contrast enhancement on the images of the developed dataset are illustrated in Fig. 4 (second column).

Once the contrast enhancement is applied by guided filter on the images, their color space is converted from RGB to $L^*a^*b^*$ to provide a better representation of acne scar regions and so achieve a better classification accuracy. In the color-opponent $L^*a^*b^*$ color space, L^* is for lightness and a^* and b^* are for the color-opponent dimensions according to the CIE XYZ color space coordinates which are compressed nonlinearly [43]. In comparison to the RGB and CMYK colour models, $L^*a^*b^*$ colour space is created to approximate human vision. L component highly matches the human perception of lightness so that it can be utilized to adjust the lightness contrast or to form the accurate color balance corrections by changing and improving the output curves in the a and b components. In contrary, RGB or CMYK spaces do not model the human visual perception and they show the output of physical devices. Hence, in these spaces, the mentioned transformations can be only realized by editing applications and applying appropriate blend modes. Consequently, in our proposed method, the color space of the images are converted from RGB to $L^*a^*b^*$. Through this conversion, a better representation is provided from the input images and the acne scar regions which significantly enhances the final classification accuracy. Some samples of $L^*a^*b^*$ color space conversion on the images of the developed dataset are illustrated in Fig. 4.

Augmentation. In deep learning-based models, one of the critical factors is providing a considerable number of training images. The original proposed five-class dataset is relatively small; thus, the neural network might face an overfitting problem during training or not be robust to different input patterns. Hence, five other data augmentation techniques, i.e., randomly rotating 30° to left or right, horizontal flipping, shearing, scaling, and translation, are applied to the training

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

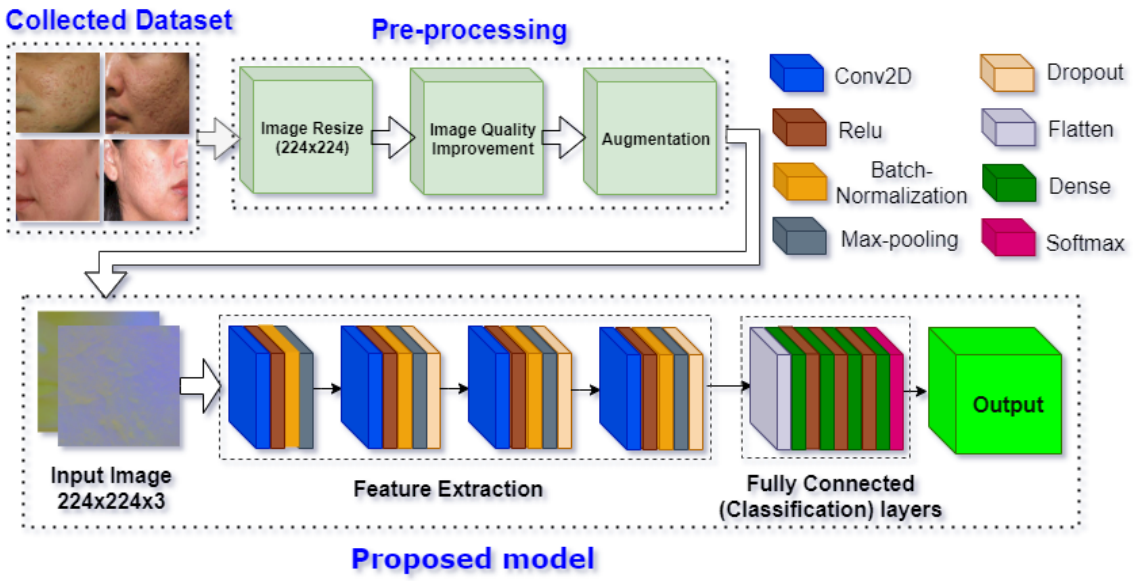

FIGURE3: The overview of the proposed *ScarNet* model for acne scar classification is composed of a pre-processing module and CNN-based feature extraction and classification.

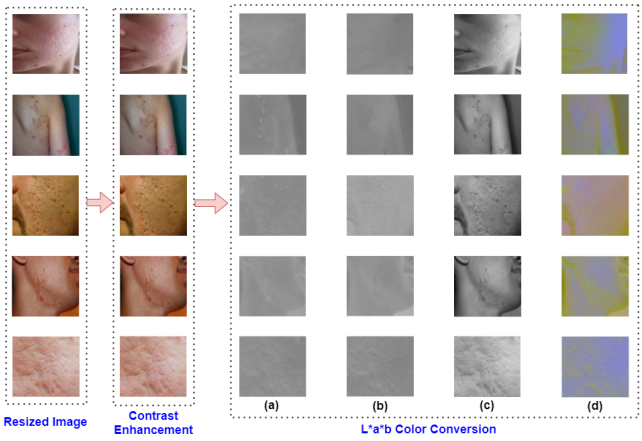

FIGURE4: Some samples of image quality improvement in the proposed *ScarNet* include contrast enhancement by guided filter and $L^*a^*b^*$ color space conversion. (a), (b), and (c), are L^*, a^*, b^*, and combinations of $L^*a^*b^*$ layers, respectively.

samples to produce additional images equal to five times the original ones. Increasing the number of training data prevents the model from overfitting and improves the performance of the system. To delve deeply into the details of the augmentation techniques, a transformation matrix is utilized to represent the augmentation methods as follows:

$$X = \begin{bmatrix} L & M & N \\ O & P & Q \\ R & S & T \end{bmatrix} \tag{2}$$

where, L and P regulate the image scaling, N and Q regulate the translation of the image, M and O regulate the shearing of the image. In detail, the rotation part of the transformation matrix is considered as:

$$X = \begin{bmatrix} cos\alpha & -sin\alpha & -cos\alpha\frac{u+1}{2} + sin\alpha\frac{v+1}{2} + \frac{u+1}{2} \\ sin\alpha & cos\alpha & -sin\alpha\frac{u+1}{2} - cos\alpha\frac{v+1}{2} + \frac{v+1}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where, u, v, and α are the picture's height, width, and rotation angle, respectively. The translation part of the transformation matrix is:

$$X = \begin{bmatrix} 1 & 0 & z_a \\ 0 & 1 & z_b \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where, z_a and z_b are the height and width for translation. The shearing is defined as:

$$X = \begin{bmatrix} 1 & -sin\mu & -j_a\frac{u+1}{2} + \frac{u+1}{2} \\ 0 & cos\mu & -\frac{u+1}{2} - cos\mu\frac{v+1}{2} + \frac{v+1}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where, μ, u, and v are the transformation intensity, the height, and the width of shearing, respectively. Scaling of the images is as:

$$X = \begin{bmatrix} j_a & 0 & -j_a\frac{u+1}{2} + \frac{u+1}{2} \\ 0 & j_b & -j_b\frac{v+1}{2} - \frac{v+1}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where, j_a and j_b regulate the zooming scales of the image.

B. MODEL ARCHITECTURE

In recent years, various CNN models have been widely used to detect and classify 2D, and 3D images [44]–[46] thanks to the existence of the significant number of data and high GPUs. CNN-based models divert scientists' attention away

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

TABLE1: The summary of the *ScarNet* architecture includes layers, filter sizes, their configurations (kernel sizes, activation functions, etc.), and the output shape of each layer.

Layer	Filter	Config.	Stride	Output
Conv2D	64	KS: 5×5; Relu	2	$224 \times 224 \times 64$
BatchNormalization	-	-	-	$224 \times 224 \times 64$
Max-pooling2D	-	KS: 2×2	2	$112 \times 112 \times 64$
Conv2D	48	KS: 3×3; Relu	2	$112 \times 112 \times 48$
BatchNormalization	-	-	-	$112 \times 112 \times 48$
Max-pooling2D	-	KS: 3×3	2	$56 \times 56 \times 48$
Dropout	-	0.4	-	$56 \times 56 \times 48$
Conv2D	32	KS: 3×3; Relu	2	$56 \times 56 \times 32$
BatchNormalization	-	-	-	$56 \times 56 \times 32$
Max-pooling2D	-	KS: 2×2;	2	$28 \times 28 \times 32$
Dropout	-	0.4	-	$28 \times 28 \times 32$
Conv2D	16	KS: 3×3; Relu	2	$28 \times 28 \times 16$
BatchNormalization	-	-	-	$28 \times 28 \times 16$
Max-pooling2D	-	KS: 2×2	2	$14 \times 14 \times 16$
Dropout	-	0.25	-	$14 \times 14 \times 16$
Flatten	-	-	-	-
Dense	-	512; Relu	-	512
Dense	-	256; Relu	-	256
Dense	-	128; Relu	-	128
Softmax	-	5	-	5

from traditional machine learning approaches in favor of higher performance, quality, and speed [47].

A novel multi-layer deep CNN model, namely *ScarNet*, is proposed for acne scar classification with minimum computational cost. It is composed of a total of 19-layers whose details are presented in Table 1. Total 15-layers are placed in four convolution blocks, and the remaining 4-layers are fully connected layers applied for classification. The input images of the first convolution block are $224 \times 224 \times 3$ images in $L^*a^*b^*$ color space. All convolution blocks contain a convolutional layer, a Batch Normalization (BN) layer, and a Max-polling (MP) layer. The applied filters in four convolutional layers are 64, 48, 32, and 16, respectively, with a kernel size (KS) of 5×5 in the first block and 3×3 in the other three blocks. As the large-sized images with a high number of pixels need more training parameters and more space size, a Max-Pooling (MP) layer is designed in all blocks. The employed KS for the MP layer in all blocks is 2×2 except in the second block, which is 3×3. In all convolutional and MP layers, zero-padding (i.e., allocating 0 values around the inputs to keep the output size equal to every kernel filter input for each convolution process) is applied with a stride of 2. BN and dropout layers are employed as regularization layers to improve the convergence, enhance the generalization, reduce the overfitting, and obtain a better performance on invisible information. BN layer is used in all four blocks while a dropout layer is applied only in each of the last three blocks setting as 0.4, 0.4, and 0.25 (i.e., 40%, 40%, and 25% of neurons in the hidden layers are considered as 0 at each training update).

The features of the input images are extracted as feature maps by passing these four blocks. The generated feature maps are fed into the last four fully connected layers, i.e., a flatten layer, three dense layers (with 512, 256, and 128 neurons, respectively), and one softmax layer, for final acne scar classification into five different classes. Rectified Linear Unit (ReLU) is used in all convolutional and dense layers as a nonlinear activation function:

$$Relu = max(0, x) \tag{7}$$

This function means that the value of 0 will be assigned if the input value is a non-positive value. Otherwise, its value remains unchanged. The last activation function of our proposed model is the softmax function which is a generalization of the logistic function to multiple classes. Through this function, the network output is normalized to a probability distribution over the predicted output classes as follows:

$$S(Z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad for \quad i = 1, 2, ..., K \tag{8}$$

where z is the input vector of K real numbers (number of classes) which is normalized into a probability distribution including K probabilities proportional to the exponentials of the input numbers. Hence, before the softmax function, the sum of vector components might not be 1 because there might be a negative or greater than one value. Once the softmax is applied, all the values are normalized in an interval of $[0, 1]$. Thus, the sum of the components of the output vector $S(Z)_i$ is guaranteed 1. In our model, the softmax function allocates distributions of probability to produce predictions for each acne scar class prediction. The training process is optimized using Adam Optimizer, which minimizes the error of the CNN model and makes it more reliable and efficient. The adopted learning rate and batch size are 0.0001 and 32.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the experimental setup, evaluation metrics, and the results are presented. The effects of the augmentation and the batch size on the final classification accuracy are also investigated. Additionally, the performance of the ScarNet is compared with four conventional ML-based classifiers and five pre-trained models on our developed dataset.

A. EXPERIMENTAL SETUP

Experiments are carried out using a 3.60 GHz Intel Core i9-10850K CPU, 64 GB RAM, and an NVIDIA GeForce RTX 2080 GPU with 8GB video memory. Pre-processing, designing the CNN architecture, and training and testing are conducted in Python, Keras, and Tensorflow platforms.

B. EVALUATION CRITERIA

To evaluate the performance of our proposed method, six different evaluation metrics are employed, i.e. accuracy ($\frac{(TP+TN)}{(TP+TN+FP+FN)}$), Recall/Sensitivity ($\frac{(TP)}{(FN+TP)}$), Precision ($\frac{(TP)}{(TP+FP)}$), F1-score ($\frac{2*(Precision*Recall)}{(Precision*Recall)}$), Kappa ($k = \frac{2\times(TP\times TN-FN\times FP)}{(TP+FP)\times(FP+TN)+(TP+FN)\times(FN+TN)}$), and Specificity ($\frac{(TN)}{(FP+TN)}$). These evaluation metrics are defined based on confusion matrix in which TP, TN, FP, and FN refer to True Positive (both real and predicted classes are 1 (True)), True Negative (both real and predicted classes are 0 (False)), False Positive (the actual class of the data

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

TABLE2: Performance comparison of the proposed *ScarNet* for five different dataset splitting scenario.

(Train - Test)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
(90 - 10) %	89.34	71.67	74.14	72.88
(80 - 20) %	**92.53**	**81.33**	**82.24**	**81.39**
(70 - 30) %	91.30	76.27	78.95	77.59
(60 - 40) %	89.67	75.01	73.77	74.38
(50 - 50) %	88.04	68.33	70.69	69.49

is 0 while the predicted one is 1), and False Negative (the actual class of the data is 1 while the predicted one is 0), respectively. The reason behind choosing False as the name is that the prediction done by the model is incorrect. To define these metrics in detail, accuracy is the number of correct predictions over the total number of predictions, also called classification rate. Recall/Sensitivity determines the correctly identified actual positive values, which is known as true positive rate. The number of predicted positive classes which belong to the actual positive class is calculated by precision. F1-score measures testing accuracy. Finally, Kappa is a rigorous method of finding the degree of agreement between two raters/judgments, categorizing N items into C categories.

C. PERFORMANCE ASSESSMENT

The performance of the proposed ScarNet is evaluated in terms of accuracy, precision, recall, specificity, and F1-Score. Images in the dataset are resized into 224×224 images whose quality is then improved by passing a guided filter and converting the color space from RGB to $L^*a^*b^*$. Splitting the pre-processed dataset into training, validation, and testing sets, five different augmentation techniques are applied to the training set to increase the dataset number and prevent the model from overfitting and make it robust for future testing data with different variations. Five different dataset splitting scenarios are investigated with 90-10%, 80-20%, 70-30%, 60-40%, and 50-50%. The first value is for training and validation sets, and the second is for the test set. The performance of our system in terms of accuracy, precision, recall, and F1-score is presented in Table 2 for all five dataset splitting. Highlighting the best performance in bold, Our proposed system ranks first for 80-20% splitting conditions in all mentioned four evaluation metrics with the values of 92.53%, 81.33%, 82.24%, and 81.39%, respectively. Hence, our experiments are based on an 80-20% splitting strategy. A total of 200 images (80%) of the dataset is used for training and validation. The rest 50 images (20%) are employed in the testing phase. It is worth mentioning that the five augmentation techniques are applied on the training set, which results in additional images equal to 5 times the original training set.

One of the critical hyperparameters required to be tuned in deep learning-based models is the batch size. It is mostly preferred by the researchers to utilize a larger batch size for training since it speeds up the computations. Adopting smaller batch sizes provides reasonable solutions with faster convergence, but there is no guarantee that the model converges to the global optima. Hence, to choose the best batch size with the number of epochs, we start our experiments at a small batch size of 16, which then increases to 32 and 64 to provide a significant trade-off between faster training dynamics and guaranteed convergence. The performance of the proposed CNN-based model is evaluated with training and validation sets for five different epoch numbers of 10, 20, 30, 40, and 50, and three different batch sizes of 16, 32, and 64. The results in terms of training accuracy (TA), validation accuracy (VA), training loss (TL), and validation loss (VL) are presented in Table 3. The highest performance is achieved for 50 epochs and the batch size of 32 with TA, VA, TL, and VL equal to 88.02%, 86.33%, 0.51%, and 0.50%, respectively.

Figure 5 (a) and (b) demonstrate the training and validation accuracy and loss, respectively, for the optimized number of epochs and batch size (i.e. 50 and 32) in our proposed *ScarNet*. The X-axis is the number of epochs in both graphs, while the Y-axis represents the accuracy or loss values. Initially, the training and validation accuracy is around 63% and 71%, respectively. Once 40 epochs pass, both graphs flatten out and reach the values of about 88% and 86%, respectively. However, there are some fluctuations in the validation accuracy before reaching its final value. The training and validation losses start from the values of about 0.66 and 0.83, respectively. Passing about 2 epochs, the validation loss dramatically decreases to 0.53. After two epochs, training and validation loss values decrease until they reach their least value of around 0.5. Considering both graphs, our proposed model performs without facing underfitting or overfitting problems.

The proposed *ScarNet* is evaluated on the test set, and the results are presented for all five acne scar classes separately in Table 4 in terms of accuracy, precision, recall, specificity, and F1-score. The highest accuracy, recall, and F1-score values are attained by Icepick with the values of 94.67%, 87.50%, and 86.67%, respectively. In two other metrics of precision and specificity, Rolling scar ranks first with the values of 88.33% and 96.86%, respectively.

D. COMPARISON WITH DIFFERENT CLASSIFIERS

To prove the capability of the classification layers of our proposed CNN-based *ScarNet*, the acne scar classification is conducted with four different machine learning classifiers and compared with the performance of our system. To follow this, the fully connected layers and the softmax layer, which are considered the classifier of our system, are replaced with machine learning-based classifiers: DT, MLP, SVM, and the Random forest. The extracted feature maps from four convolution blocks of *ScarNet* are fed into these classifiers. The performance of these classifiers is evaluated and compared with our classification system based on the softmax classifier in terms of accuracy, precision, sensitivity, and specificity in Table 5. Our model outperforms all four classifiers in terms of all evaluation metrics with the values of 92.53% (accuracy), 81.33% (precision), 82.34% (sensitivity), and 95.38%

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

TABLE3: The performance of the *ScarNet* is shown for five epoch numbers with three batch sizes in terms of TA (training accuracy), VA (validation accuracy), TL (training loss), and VL (validation loss).

	Batch Size 16				Batch Size 32				Batch Size 64			
Epoch	TA (%)	VA (%)	TL	VL	TA (%)	VA (%)	TL	VL	TA (%)	VA (%)	TL	VL
10	62.77	71.61	1.23	1.82	82.7	82.1	0.54	0.48	69.12	59.89	1.59	1.32
20	69.53	74.38	1.02	1.35	82.31	86.23	0.53	0.49	71.91	61.87	1.44	1.23
30	75.64	77.51	0.87	1.12	86.87	85.07	0.53	0.48	73.15	62.74	1.3	1.08
40	78.44	79.37	0.75	0.98	87.54	85.61	0.52	0.47	76.56	64.17	1.15	0.82
50	80.44	83.51	0.68	0.91	**88.02**	**86.33**	**0.51**	**0.50**	22.67	67.65	0.91	0.61

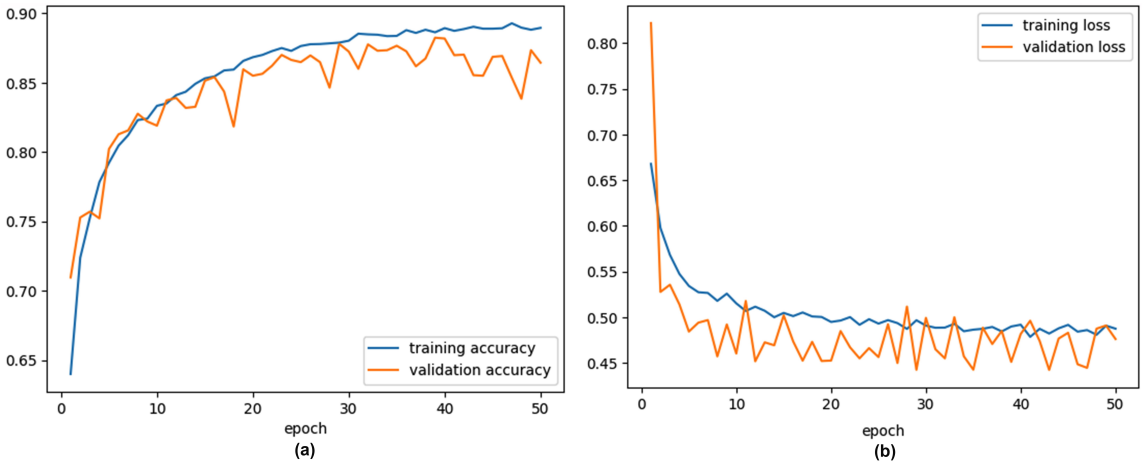

FIGURE5: Training and validation (a) accuracy and (b) loss graphs of the proposed *ScarNet*.

TABLE4: The performance of the *ScarNet* for each scar class and their average values in terms of accuracy (Acc), precision (Pre), recall (Rec), specificity (Spe), and F1-Score.

	Acc (%)	Pre (%)	Rec (%)	Spe (%)	F1-Score (%)
Boxcar	93.00	80.00	84.21	95.06	82.05
Hypertrophic	91.33	70.00	84.00	92.80	77.37
Ice Pick	**94.67**	86.67	**87.50**	96.67	**86.67**
Keloid	94.00	81.67	86.67	95.49	84.48
Rolling	89.67	**88.33**	68.83	**96.86**	76.36
Average	92.53	81.33	82.24	95.38	81.38

TABLE5: Performance comparison of the proposed *ScarNet* is shown with different machine learning-based classifiers.

Classifiers	Accuracy	Precision	Sensitivity	Specificity
DT	84.97%	74.23%	73.07%	84.69%
MLP	85.44%	74.32%	73.91%	85.58%
SVM	87.93%	76.11%	74.96%	87.88%
Random Forest	89.17%	78.41%	77.59%	91.85%
ScarNet	**92.53%**	**81.33%**	**82.34%**	**95.38%**

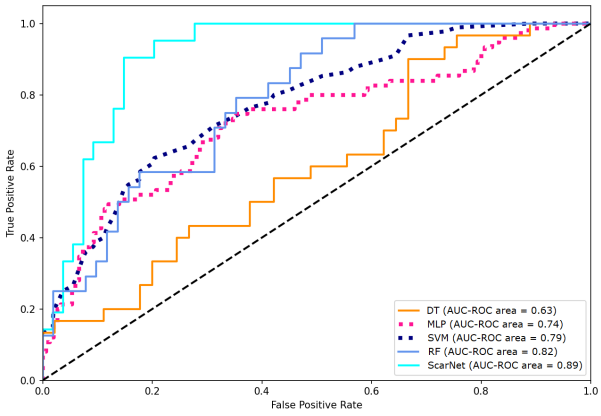

FIGURE6: The AUC of the ROC curves belonging to the *ScarNet* and four ML-based classifiers is shown. Each line indicates the performance of each classifier.

(specificity). Comparing the accuracy of DT (84.97%), MLP (85.44%), SVM (87.93%), and random forest (89.17%), our proposed *ScarNet* achieves at least 3.36% more accuracy than others.

As another evaluation tool, the performance of our system is compared with four classifiers based on the Receiver Operating Characteristic (ROC) Curve as illustrated in Fig. 6. It illustrates the trade-off between the true positive rate (TPR) as the y-axis and the false positive rate (FPR) as the x-axis at different thresholds. The AUC (area under the curve) indicates the classification capability in this probability curve by determining whether the curve is above or below the diagonal line (chance level). The AUC scores of the DT, MLP, SVM, RF, and the proposed *ScarNet* are 0.63, 0.74, 0.79, 0.82, and 0.89, respectively. Since a larger AUC indicates the greater performance, our proposed *ScarNet* surpasses the others with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

TABLE6: Performance comparison with and without data augmentation between the proposed *ScarNet* and five pre-trained models on the developed dataset is shown.

Dataset	Models	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	Kappa (%)
Without (augmentation)	Inception-v3 [48]	75.61	47.12	48.48	48.33	47.01
	MobileNet [49]	76.72	50.41	49.90	50.28	47.58
	VGG-16 [50]	78.77	51.82	50.39	51.08	48.69
	ResNet-50 [51]	80.27	52.44	51.50	52.94	49.75
	AlexNet [52]	81.34	53.41	53.37	54.08	51.28
	ScarNet	**81.89**	**54.02**	**53.46**	**54.79**	**53.60**
With (augmentation)	Inception-v3 [48]	87.43	76.52	76.11	75.05	73.23
	MobileNet [49]	88.21	77.56	76.68	77.31	73.80
	VGG-16 [50]	89.31	80.65	77.03	78.97	74.14
	ResNet-50 [51]	89.88	79.67	77.5	79.36	74.01
	AlexNet [52]	91.47	78.63	78.84	82.23	74.81
	ScarNet	**92.53**	**81.33**	**82.24**	**95.38**	**76.70**

TABLE7: Comparison between the proposed *ScarNet* and five pre-trained models in terms of network parameters and accuracy.

Parameters	Inception-v3 [48]	MobileNet [49]	ResNet-50 [51]	AlexNet [52]	VGG-16 [50]	*ScarNet*
Model Size	92MB	49MB	98MB	294MB	528MB	18MB
Training Parameters	23.9M	4.2M	25.6M	61M	138M	**0.895M**
No. of Layers	48	30	177	25	41	**19**
Epochs	200	200	200	200	200	**50**
Running-time (s)	2535	2262	1785	1092	2685	**217**
Accuracy	87.43%	88.21%	89.88%	91.47%	89.31%	**92.53%**

a maximum AUC of 0.89.

E. COMPARISON WITH PRE-TRAINED MODELS

To further prove the capability and efficiency of our proposed *ScarNet*, it is compared with five pre-trained models, i.e. Inception-v3 [48], MobileNet [49], VGG-16 [50], ResNet-50 [51], and AlexNet [52], on our proposed dataset in Table 6. Considering two scenarios for the dataset, i.e., training set with and without applying augmentation, the impact of the augmentation on the system's performance is also investigated. The results are compared in terms of accuracy, precision, recall, specificity, and kappa score. First, comparing the results of all models in two dataset scenarios, it is observed that the accuracy is increased at least 10% after applying the augmentation. The performance of our proposed *ScarNet* before augmentation is 81.89%, 54.02%, 53.46%, 54.79%, and 53.60% in terms of accuracy, precision, recall, specificity, kappa score, respectively. After augmentation, these values are significantly increased to 92.53%, 81.33%, 82.24%, 95.38%, and 76.70%, respectively. Similar improvements are observed in the other pre-trained models as well due to augmentation. Hence, it is concluded that the augmentation can extremely improve the classification accuracy and prevent the model from overfitting. It also makes the system robust for the possible various future test data. Moreover, the proposed system outperforms the other five pre-trained models in all evaluation metrics in both with and without augmentation scenarios.

Our proposed model is also compared with the other pre-trained models in terms of network characteristics such as model size, training parameters, number of layers, number of epochs, and running time is Table 7. Our *ScarNet* not only achieves the highest classification accuracy but also has the most diminutive model size and the training parameters (18 MB and 0.89M, respectively). The largest model size belongs to VGG-16, with a value of 528MB. Our model achieved competitive performance by providing only 19 layers in 50 epochs (1/4 of the other models' epochs). In contrast, the number of layers increases even up to 177 in ResNet-50. The running time of our model is 217s which is at least five times less than those of other pre-trained models. These outstanding values for the network parameters of our CNN-based model and its superior performance prove the effectiveness and efficiency of our system in automatic acne scar classification.

F. FAILURE CASES

A graphical representation of acne scar classification based on our *ScarNet* is illustrated in Fig. 7. Observing the correctly classified cases, our model can successfully categorize the acne scars into five classes with a high score difference. Despite the high-performance accuracy of our system, there are also a few misclassification cases, as demonstrated in the second row of Fig. 7 (actual and wrongly predicted labels are represented with blue and red colors, respectively). However, comparing the confidence score, *ScarNet* classifies the acne scars wrongly, mostly with a slight difference in scores. We draw inferences from deeply analyzing the misclassified acne images. The proposed system mostly challenges classifying the images with shadows, highly bright illumination, poor clarity, or low image quality. Misclassification also occurs on different acne scars with small skin regions simultaneously (e.g., fourth example in the misclassification cases).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138021, IEEE Access

FIGURE7: Graphical representation of acne scar classification. The top row is the correctly classified cases. The bottom row is the samples of wrongly classified images (blue and red colors indicate the actual and incorrectly predicted classes, respectively).

VI. CONCLUSION AND FUTURE WORKS

This paper presented an automated end-to-end CNN-based network for acne scar classification. An acne scar dataset was developed by capturing and collecting 250 images in 5 classes. These images were resized, and their quality was improved by designing a pre-processing module. The number of images was increased by applying five different augmentation techniques, which prevented the model from overfitting, increased its performance, and made it more robust. A novel deep CNN-based model was proposed to extract the feature maps from these images based on four convolution blocks and then classify them into five classes using fully connected and softmax layers. The performance of this system was investigated for different dataset splitting scenarios, number of epochs, and batch sizes. In this method, the influence of the augmentation was also analyzed. The feasibility and competitive performance of the proposed *ScarNet* was proved by conducting an extensive comparison with four conventional ML-based classifiers and five pre-trained models. Our system could accurately classify the acne scars with minimized computational cost compared to the other pre-trained models by adjusting the optimizer, loss function, activation functions, filter and kernel sizes, regularization methods, and batch size. Experimental results demonstrated that the accuracy of our model is increased by 1.06%, 2.65%, 3.22%, 4.32%, and 5.1% compared to AlexNet, ResNet-50, VGG-16, MobileNet, and Inception-v3, respectively, without compromising the computational cost. As our future research work, we plan to extend our proposed acne scar dataset with more images and more classes for detecting the depth (shallow or deep) of some types of acne scars. Additionally, we will provide an automated system to segment and annotate different types of acne scars that appear in a close adjacent to each other on a small skin region.

REFERENCES

[1] G. Fabbrocini, M. Annunziata, V. D'arco, V. De Vita, G. Lodi, M. Mauriello, F. Pastore, and G. Monfrecola, "Acne scars: pathogenesis, classification and treatment," Dermatology research and practice, vol. 2010, 2010.

[2] J. Koo, "The psychosocial impact of acne: patients' perceptions," Journal of the American Academy of Dermatology, vol. 32, no. 5, pp. S26–S30, 1995.

[3] B. Dreno and F. Poli, "Epidemiology of acne," Dermatology, vol. 206, no. 1, p. 7, 2003.

[4] J. K. Tan and K. Bhate, "A global perspective on the epidemiology of acne," British Journal of Dermatology, vol. 172, pp. 3–12, 2015.

[5] S. Z. Ghodsi, H. Orawa, and C. C. Zouboulis, "Prevalence, severity, and severity risk factors of acne in high school pupils: a community-based study," Journal of investigative Dermatology, vol. 129, no. 9, pp. 2136–2141, 2009.

[6] A. Börve, "How long do you have to wait for a dermatologist appointment?" in https://www.firstderm.com/appointment-wait-time-see-dermatologist/, 2021, [Online; accessed 14-April-2021].

[7] C. I. Jacob, J. S. Dover, and M. S. Kaminer, "Acne scarring: a classification system and review of treatment options," Journal of the American Academy of Dermatology, vol. 45, no. 1, pp. 109–117, 2001.

[8] A. Goodarzi, E. Behrangi, M. Ghassemi, N. N. Nobari, A. Sadeghzadeh-Bazargan, and M. Roohaninasab, "Acne scar; a review of classification and treatment," J Crit Rev, vol. 7, no. 5, pp. 815–823, 2020.

[9] D. Fife, "Practical evaluation and management of atrophic acne scars: tips for the general dermatologist," The Journal of clinical and aesthetic dermatology, vol. 4, no. 8, p. 50, 2011.

[10] G. J. Goodman and J. A. Baron, "Postacne scarring–a quantitative global scarring grading system," Journal of cosmetic dermatology, vol. 5, no. 1, pp. 48–52, 2006.

[11] B. Dreno, A. Khammari, N. Orain, C. Noray, C. Merial-Kieny, S. Méry, and T. Nocera, "Ecca grading scale: an original validated acne scar grading scale for clinical practice in dermatology," Dermatology, vol. 214, no. 1, pp. 46–51, 2007.

[12] P. K. Upadhyay and S. Chandra, "An improved bag of dense features for skin lesion recognition," Journal of King Saud University-Computer and Information Sciences, 2019.

[13] Z. Wu, S. Zhao, Y. Peng, X. He, X. Zhao, K. Huang, X. Wu, W. Fan, F. Li, M. Chen et al., "Studies on different cnn algorithms for face skin disease classification based on clinical images," IEEE Access, vol. 7, pp. 66 505–66 511, 2019.

[14] P. Tang, Q. Liang, X. Yan, S. Xiang, and D. Zhang, "Gp-cnn-dtel: Global-part cnn model with data-transformed ensemble learning for skin lesion classification," IEEE journal of biomedical and health informatics, vol. 24, no. 10, pp. 2870–2882, 2020.

[15] R. Patil and S. Bellary, "Machine learning approach in melanoma cancer stage detection," Journal of King Saud University-Computer and Information Sciences, 2020.

[16] M. Combalia, F. Hueto, S. Puig, J. Malvehy, and V. Vilaplana, "Uncertainty estimation in deep neural networks for dermoscopic image classification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 744–745.

[17] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. de Oliveira Marinho, J. Gallegos, S. Gabriele et al., "A deep learning system for differential diagnosis of skin diseases," Nature Medicine, vol. 26, no. 6, pp. 900–908, 2020.

[18] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, "Deep learning in skin disease image recognition: A review," IEEE Access, 2020.

[19] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative feature learning for skin disease classification using deep convolutional neural network," IEEE Access, vol. 8, pp. 39 025–39 033, 2020.

[20] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," Computerized Medical Imaging and Graphics, vol. 88, p. 101843, 2021.

[21] R. Mohakud and R. Dash, "Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection," Journal of King Saud University-Computer and Information Sciences, 2021.

[22] R. Ramli, A. S. Malik, A. F. M. Hani, and F. B.-B. Yap, "Identification of acne lesions, scars and normal skin for acne vulgaris cases," in 2011 National Postgraduate Conference. IEEE, 2011, pp. 1–4.

[23] B. C. Dey, B. Nirmal, and R. R. Galigekere, "Automatic detection of acne scars: Preliminary results," in 2013 IEEE Point-of-Care Healthcare Technologies (PHT). IEEE, 2013, pp. 224–227.

[24] N. Alamdari, K. Tavakolian, M. Alhashim, and 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.

[25] X. Shen, J. Zhang, C. Yan, and H. Zhou, "An automatic diagnosis method of facial acne vulgaris based on convolutional neural network," Scientific reports, vol. 8, no. 1, pp. 1–10, 2018.

[26] M. S. Junayed, A. A. Jeny, S. T. Atik, N. Neehal, A. Karim, S. Azam, and 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.

[27] A. Z. Hameed, W. K. Awad, N. A. Irsan, and A. S. Abdulbaqi, "Hybrid technique for skin pimples image detection and classification," International Journal of Advanced Science and Technology, vol. 29, no. 3, pp. 4102–4109, 2019.

[28] X. Wu, N. Wen, J. Liang, Y.-K. Lai, D. She, M.-M. Cheng, and J. Yang, "Joint acne image grading and counting via label distribution learning," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10 642–10 651.

[29] K. Rashataprucksa, C. Chuangchaichatchavarn, S. Triukose, S. Nitinawarat, M. Pongprutthipan, and K. Piromsopa, "Acne detection with deep neural networks," in 2020 2nd International Conference on Image Processing and Machine Vision, 2020, pp. 53–56.

[30] C.-Y. Chang, H.-Y. Liao et al., "Automatic facial spots and acnes detection system," Journal of Cosmetics, Dermatological Sciences and Applications, vol. 3, no. 01, p. 28, 2013.

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

[32] M. Khongsuwan, S. Kiattisin, W. Wongseree, and A. Leelasantitham, "Counting number of points for acne vulgaris using uv fluorescence and image processing," in The 4th 2011 Biomedical Engineering International Conference, 2012, pp. 142–146.

[33] F. S. Abas, B. Kaffenberger, J. Bikowski, and 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.

[34] N. Kittigul and B. Uyyanonvara, "Acne detection using speeded up robust features and quantification using k-nearest neighbors algorithm," in Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science, 2017, pp. 168–171.

[35] T. Zhao, H. Zhang, and J. Spoelstra, "A computer vision application for assessing facial acne severity from selfie images," arXiv preprint arXiv:1907.07901, 2019.

[36] C. Kulshrestha, P. Juyal, S. Sharma, and S. Juyal, "Machine learning algorithms based skin disease detection," vol. 9, pp. 4044–4049, 05 2020.

[37] M. S. Junayed, A. N. M. Sakib, N. Anjum, M. B. Islam, and 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.

[38] G. J. Goodman, "Postacne scarring: a review of its pathophysiology and treatment," Dermatologic surgery, vol. 26, no. 9, pp. 857–871, 2000.

[39] L. A. Zaleski-Larsen, S. G. Fabi, T. McGraw, and M. Taylor, "Acne scar treatment: a multimodality approach tailored to scar type," Dermatologic Surgery, vol. 42, pp. S139–S149, 2016.

[40] A. Skin, "Dermnet nz," https://dermnetnz.org/, 2021, [Online; accessed 15-October-2020].

[41] K. He, J. Sun, and X. Tang, "Guided image filtering," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 6, pp. 1397–1409, 2012.

[42] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Sixth international conference on computer vision (IEEE Cat. No. 98CH36271). IEEE, 1998, pp. 839–846.

[43] P. J. Baldevbhai and R. Anand, "Color image segmentation for medical images using l* a* b* color space," IOSR Journal of Electronics and Communication Engineering, vol. 1, no. 2, pp. 24–45, 2012.

[44] M. S. Junayed, N. Anjum, A. Noman, and B. Islam, "A deep cnn model for skin cancer detection and classification," 2021.

[45] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," Health information science and systems, vol. 8, no. 1, pp. 1–8, 2020.

[46] S. Bagchi, A. Banerjee, and D. R. Bathula, "Learning a meta-ensemble technique for skin lesion classification and novel class detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 746–747.

[47] I. Fantini, C. Yasuda, M. Bento, L. Rittner, F. Cendes, and R. Lotufo, "Automatic mr image quality evaluation using a deep cnn: A reference-free method to rate motion artifacts in neuroimaging," Computerized Medical Imaging and Graphics, vol. 90, p. 101897, 2021.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[49] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.