A cell phone app for facial acne severity assessment

Jiaoju Wang[1] · Yan Luo[2] · Zheng Wang[1,3] · Alphonse Houssou Hounye[1] · Cong Cao[1] · Muzhou Hou[1] ● · Jianglin Zhang[4,5]

Abstract

Acne vulgaris, the most common skin disease, can cause substantial economic and psychological impacts to the people it affects, and its accurate grading plays a crucial role in the treatment of patients. In this paper, we firstly proposed an acne grading criterion that considers lesion classifications and a metric for producing accurate severity ratings. Due to similar appearance of acne lesions with comparable severities and difficult-to-count lesions, severity assessment is a challenging task. We cropped facial skin images of several lesion patches and then addressed the acne lesion with a lightweight acne regular network (Acne-RegNet). Acne-RegNet was built by using a median filter and histogram equalization to improve image quality, a channel attention mechanism to boost the representational power of network, a region-based focal loss to handle classification imbalances and a model pruning and feature-based knowledge distillation to reduce model size. After the application of Acne-RegNet, the severity score is calculated, and the acne grading is further optimized by the metadata of the patients. The entire acne assessment procedure was deployed to a mobile device, and a phone app was designed. Compared with state-of-the-art lightweight models, the proposed Acne-RegNet significantly improves the accuracy of lesion classifications. The acne app demonstrated promising results in severity assessments (accuracy: 94.56%) and showed a dermatologist-level diagnosis on the internal clinical dataset.The proposed acne app could be a useful adjunct to assess acne severity in clinical practice and it enables anyone with a smartphone to immediately assess acne, anywhere and anytime.

Keywords Deep learning(DL) · Acne vulgaris · Cell phone app · Severity assessment · Acne regular network

1 Introduction

Automatic assessments of skin disease severity is important in the medical field. Acne vulgaris is a common chronic inflammatory skin disease involving pilosebaceous units characterized by skin with seborrhea, comedones, papules, nodules, pimples and scarring [1, 2]. It has affected up to 650 million people and 85% of the adolescents and young adults in the world [3]. As a damaging skin disease, acne not only causes physical harm to patients but also has a significant impact on a patient psychology. Acne often causes low self-esteem, depression, anxiety, and even suicidal tendencies [4, 5]. As a result, there is a significant demand for timely and effective acne diagnosis and management. Assessing the severity of the acne is vital for dermatologists to make a precise standardized treatment decision [6]. The severity standard criteria commonly used by dermatologists is the Global Acne Grading System (GAGS) [7] and Hayashi criteria [6]. Specifically, acne can be graded into four levels of severity, mild, moderate, severe, and very severe. The traditional assessment method is to observe and assess acne severity through a dermatologist in the hospital, which is time-consuming and is often dependent on the expertise of the dermatologist [8]. Moreover, patients also face difficulties in hospital registration, long hospital waiting times and inconvenient transportation, especially for patients in remote and underdeveloped areas. An alternative solution to resolve these issues is consulting a dermatologist by sending images or videos but these electronic transmission methods may cause privacy-related problems and are time-consuming for dermatologists. Therefore, a reliable and convenient computer-aided diagnosis method for acne is imperative for effective and efficient treatment.

Jiaoju Wang and Yan Luo contributed equally to this work.

✉ Muzhou Hou
  houmuzhou@sina.com

Extended author information available on the last page of the article.

In recent years, substantial advances have been made in acne analysis [9]. Due to the excellent performance of deep learning(DL) in image processing tasks, convolutional neural network (CNN)-based techniques are finding ever greater use in acne lesion analysis [8–11]. Seit et al. developed an artificial intelligence algorithm to determine the severity of facial acne and identify different types of acne lesions (comedonal, inflammatory) and post inflammatory hyperpigmentation (PIHP) or residual hyperpigmentation [12]. Compared with most methods relying on handcrafted features, CNN-based methods have a higher performance and are more flexible. However, there are still some challenging problems when employing CNNs for the assessment of acne severity. First, the commonly used acne grading criteria are based solely on lesion type or lesion quantity. This is not a well thought out process that may cause CNN-based methods to not capture accurate grading results. Second, acne images with comparable severities have similar appearances and are difficult to distinguish. A commonly used solution for this issue is adopting a wider and deeper network, which requires large amounts of computational resources (more powerful GPUs) but cannot be deployed in low-specification environments, such as mobile phones. Additionally, previous works have focused on diagnoses using only acne images instead of integrating patient information for a more personalized diagnosis. Thus, automatic personalized assessments of acne severities, which can be acquired without regard to time and location, is a challenging task.

In this article, we address acne lesion classification with a lightweight deep learning architecture called the Acne regular network (Acne-RegNet) and we also propose a cell phone app for facial acne grading based on the new severity grading criteria. Combining the Global Acne Grading System (GAGS) and the Hayashi criteria, new severity grading criteria are proposed, which incorporate the outcome measures of lesion counting and classification. The facial photos are first cropped for some lesion patches. Acne-Regnet is built for patch classification (comedones, pimples, pustules, nodules and normal) based on a regular network (RegNet) [13], which is a fast network using residual bottlenecks with group convolutions. We then modified it by using (i) a median filter and contrast-limited adaptive histogram equalization (CLAHE) to handle the noise and color variations of the acne images, (ii) a lightweight attention mechanism to improve the representational power of a network, (iii) a region-based focal loss to alleviate the problem of class imbalance, and (iv) model pruning and feature-based knowledge distillation to reduce the model parameters. On this basis, the metadata of patients are used for personalized diagnosis, and the acne severity grading of the patients can be obtained. The smartphone application for acne assessment is also designed. With this technology, users who are concerned about facial acne can download the app, take a photo of his/her face, draw the detection area, and instantaneously receive acne severity results and recommended medications.

The major research contributions in this study consist of the following:

- We propose a novel facial acne severity grading method to achieve a more accurate computer-aided diagnosis.
- We propose a lightweight Acne-RegNet model using a channel attention module, region-based focal loss, model pruning and knowledge distillation to provide a good trade-off between computational memory and performance in the classification of acne lesions.
- To attain personalized diagnoses, instead of only considering images, we leverage metadata such as demographic information, which improves the assessment performance.
- We design and implement a mobile phone application with an automatic acne grading method and a user-friendly interface.
- On internal clinical experiment, the acne app demonstrates a comparable performance with actual dermatologists and has stable results with different background lighting conditions, devices and skin tones.

The manuscript is organized as follows. Section 1 describes the introduction of the acne analysis and their evolution using deep learning. Section 2 gives information about the related work presented in the literature. Section 3 describes proposed methodology. Section 4 describes dataset,experimental details and evaluation metrics. Results and discussion have been shown in Section 5. Section 6 concludes the work.

2 Related works

Due to the excellent feature extraction abilities of convolutional neural networks (CNNs), deep learning has been increasingly used in the field of medical disease diagnosis in recent years [14–16], in fields such as dermatology [17–20], ophthalmology [21, 22], and radiology [23, 24]. Shen et al. applied fluorescence imaging and deep convolutional neural networks to diagnose intraoperative gliomas in real time for the rapid and accurate identification of gliomas during surgery [25]. Another study used functional magnetic resonance imaging (FMRI) data to construct a kernel-based CNN framework for identifying hierarchical features for brain disease diagnoses [26]. In a cross-sectional experiment, a high-performance CNN trained on a large-scale 12-lead ECG diagnosis outperformed the clinical diagnosis of a cardiologist for most classes [27]. Chaddard et al. proposed encoding feature distributions learned from CNNs using

Gaussian mixture models (GMMs) to improve COVID-19 predictions in chest CT and X-ray scans [27]. Their research shows the excellent feature representation abilities of deep networks for medical images. However, most CNN-based models require a large quantity of computational resources and are difficult to deploy on low-specification devices, such as cell phones.

To overcome this limitation, several lightweight models have been developed in recent years. MobileNets, proposed by Google for embedded devices such as mobile phones, are a series of lightweight models that use depthwise separable convolution and several techniques to reduce the number of parameters [28, 29]. ShuffleNet achieves high accuracy with fewer parameters by classifying feature lines using groupwise convolutions prior to convolution to provide a smooth information flow [30]. Based on depthwise separable convolutions and group wise convolutions, Kaddar et al. designed an efficient mobile network called a diversity network using a hierarchical architecture, [31]. Hu et al. proposed squeeze-and-excitation networks (SENets), which extensively uses SE blocks to reduce the number of parameters and to reduce the number of model parameters and the computational complexity in convolutional layers [32]. EffcientNet was proposed using a scaling method that uniformly scales all dimensions of depth, width and resolution using a simple yet highly effective compound coefficient [33]. Since traditional convolution layers often extract many similar features, a Ghost module was introduced by Han et al. to generate more feature maps from cheap operations. Then, Ghost bottlenecks were designed, and a lightweight GhostNet was finally established to effectively reduce the computational cost [30]. Passalis et al. combined BoF-based pooling layers with CNNs to train a lightweight deep CNN using feature packet pooling and to allow end-to-end supervised training using the backpropagation algorithm [34]. Anamorphic network(Anam-Net) is a lightweight cellular neural network based on deformable deep embedding that can segment abnormal regions in chest computed tomography (CT) images of COVID-19. The model size of Anam-Net is approximately tens of megabytes and can be easily deployed on mobile platforms [35]. These mobile neural networks achieve promising performances under a given limited computational scenario using efficient modules such as SE blocks, and it is anticipated that they will be deployed on embedded devices. However, these lightweight architectures often do not perform as well as the general deep network models, and thus designing small and efficient mobile neural networks is a challenging task.

For a given neural network, model compression may be another effective method to reduce the parameters and computation. Researchers have proposed some model pruning techniques to reduce the model size with an allowed accuracy range by removing useless channels for easier acceleration or unimportant connections between neurons [36, 37]. In deep neural networks, weights are usually stored in the form of 32-bit floating-point numbers [38], and the model quantization compresses the original network by reducing the number of bits required to represent each weight [28, 39]. To achieve the trade-off between computational expense and performance, knowledge distillation (KD) has been proposed to transfer the information learned from a larger model with better performance to a lightweight small model [40–42]. These technologies are attracting increasing attention as they are often efficient on both CPUs and GPUs without requiring special implementation or losing the accuracy of the given trained model.

With the development of phone hardware technologies and the widespread availability of smartphones, medical applications have become increasingly popular with patients for disease management and intervention, reducing the burden of attending face-to-face medical visits. Duarte-Rojo et al. developed a smartphone app for cirrhosis patients, with educational and exercise videos of an appropriate intensity, to provide assisted rehabilitation and remote supervision [43]. Krishnamurti et al designed an antenatal care smartphone app to enhance the identification of patients at risk for preeclampsia [44]. Mobile applications have also been increasingly important in dermatology in recent years [45]. However, most studies generated auxiliary diagnoses by transmitting photos to a server, and only a few studies created mobile applications to diagnose skin diseases from photographs using lightweight models [45]. Goceri proposed a mobile technology by a modified MobileNet for the classification of five widespread skin diseases: hemangioma, acne vulgaris, psoriasis, rosacea and seborrheic dermatitis [46]. Verma et al. designed a mobile app using hybrid deep neural networks integrating a multilayer perceptron (MLP) network and NasNetMobile into a single framework for the diagnosis of hand, foot and mouth disease using integrated features from clinical and image data [47]. Thus, a cell phone app for acne assessment would be useful and is desperately needed.

3 Proposed methodology

The entire procedure of our proposed scheme is demonstrated in Fig. 1. The segmentation model is first used to recognize the different facial regions, such as the nose, cheeks, forehead, and chin, from the Xiangya Hospital and ACNE04 datasets. Each region is cropped into several patches, which are the inputs to the Acne-RegNet model for automatic recognition. The trained Acne-RegNet model can be deployed to mobile phones and used to design an app

Fig. 1 Flowchart of the proposed auxiliary diagnosis app based on the Acne-RegNet model
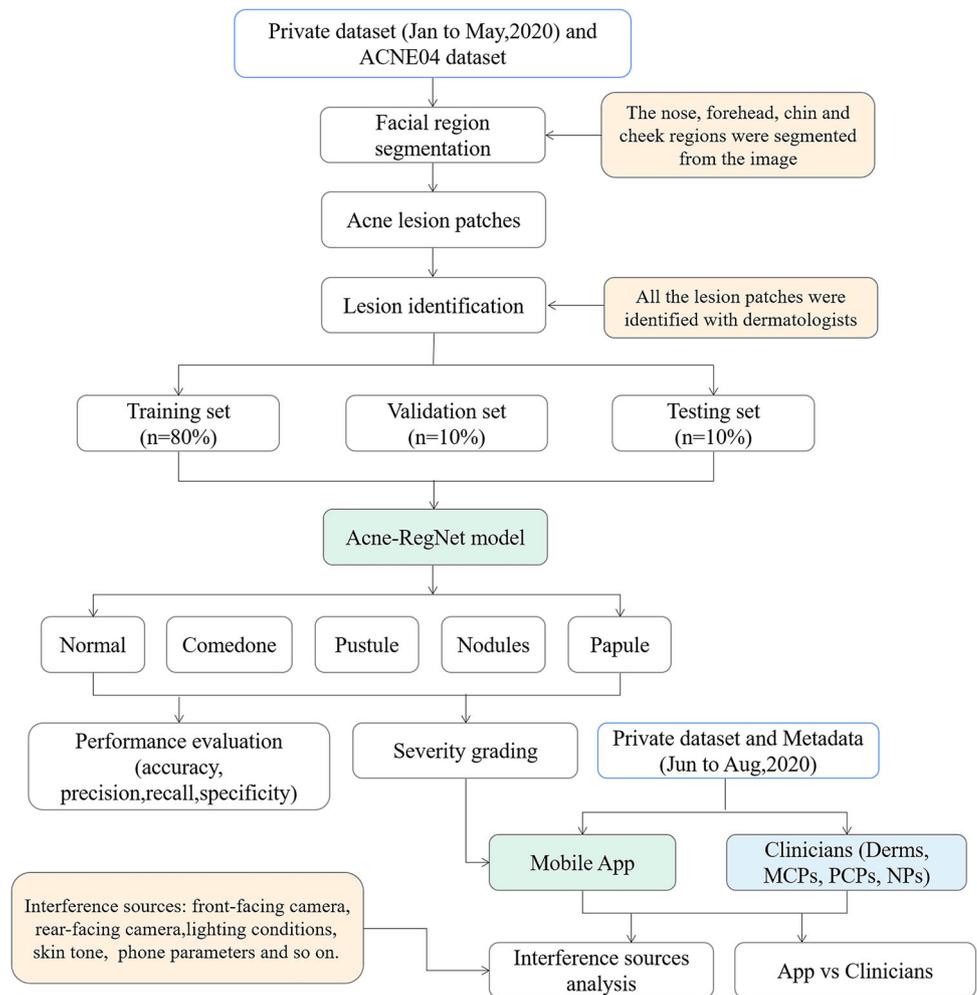
for facial acne severity grading. The model architectures, training process, mobile app and measurement methods are discussed in the subsequent sections.

3.1 Severity criteria

In this study, we combined the Global Acne Grading System (GAGS) [7] and Hayashi criterion [6] to establish the acne grading criterion. It is an evidence-based comprehensive grading criterion and is expected to yield consent from most dermatologists. The proposed severity criterion considers four regions (chin, cheeks, nose and forehead) on the face and the lesion classification and conducts a count at each location, resulting in a severity score for a severity rating. Specifically, we count the number of various types of lesions in each region separately and assign scores to various lesions according to the GAGS criteria (comedone:

1, papule: 2, pustule: 3, nodule: 4). Then, the severity score($Score$) is calculated by:

$$Score = \sum_{i=1}^{4} r_i * (I_{comedone} * 1 + N_{papule} * 2 + N_{pustule} * 3 \\ + N_{nodule} * 4) \tag{1}$$

Where $I_{comedone}$ is the indicator function. Since comedones are often numerous and difficult to detect, we take no account of their number when calculating the severity score. N_{papule}, $N_{pustule}$ and N_{nodule} denote the number of papules, pustules and nodules in each region, respectively. r_i represents the regional factors (forehead: 2, nose: 1, cheek: 2, chin: 1), which are based roughly on the surface area, distribution and density of pilosebaceous units. Moreover, five independent expert dermatologists graded the collected

facial photographs relying on their clinical experience. The results of the rating from these dermatologists is used as a severity grading for each patient. We then analyzed the relationships between the severity rating and severity score and obtain the rating boundaries. The proposed severity criteria are also verified in the test set. The details and corresponding treatment recommendations [48] are shown in Table 1.

3.2 Acne assessment method

According to the severity criteria, the core of an automatic acne assessment is the facial region distinction and lesion classification. In this work, we use a segmentation model called the pyramid scene parsing network (PSPNet) to obtain facial skin images in different locations and propose a classification model (Acne-RegNet) to automatically recognize the lesions of each region.

3.2.1 Facial region segmentation

Since our work focuses on the face of each patient and requires detecting the forehead, chin, nose and cheek of the face, we try to use a PSPNet to automatically obtain the target region. PSPNet is a commonly used segmentation model and is capable of global context information by a different-region-based context aggregation using a pyramid pooling module [49]. To develop architectures that optimize the accuracy latency trade-off on low-specification devices, we adopt MobibleNet v3 as the backbone of PSPNet. MobileNetv3 is a popular lightweight CNN architecture with depthwise separable convolutions, an inverted residual with linear bottlenecks and squeeze and excitation blocks [29]. In this way, we can efficiently implement facial region segmentations with a lower computational expense.

3.2.2 Acne regular network

For the acquired facial skin images, we then meshed them into a set of 80×80 acne lesion patches (normal, comedone, pimple, pustule, nodule) as the input to the proposed lesion classification model. Figure 2 shows the overview of our proposed model for acne lesion classification, which is the main part of our app algorithm . The contrast limited histogram equalization method is first adopted to enhance the patch quality and improve the lesion features. The processed lesion patches were then resized to 224×224 by bilinear interpolation for input to the modified regular network.

In detail, we format the pipeline for the classification architecture as a combination of three key parts, the stem, the body part (four stages) and the head. The stem network is a 3×3 convolution (conv) with a stride of 2. The

Table 1 Severity criteria and Treatment strategy

Severity criteria and treatment strategy	Severity rating			
	Mild	Moderate	Severe	Very severe
Severity criteria	$1 < score \leq 40$	$40 < score \leq 90$	$90 < score \leq 160$	$score > 160$
Treatment strategy	Topical retinoid or Fixed combination with retinoid or benzoyl peroxide or Salicylic acid or Azelaic acid	Fixed combination or benzoyl peroxide or Topical Retinoid or Azelaic Acid	Fixed combination Preferred + Hormonal therapy or Oral Antibiotic	Fixed combination + Oral Antibiotic Preferred or Oral Isotretinoin or Oral Hormonal therapy

network body conducts the bulk of the computation and contains four stages operating at a progressively reduced resolution. Each stage contains a sequence of identical blocks (Fig. 2(Stage)). The block is based on a standard residual block with a group of convolutions. Each block consists of 1×1 conv, followed by a 3×3 group conv and a squeeze-and-excitation (SE) block [32] and a final 1×1 conv. Moreover, BatchNorm (BN) and ReLU followed each conv. In particular, the block parameters (block width b_i, number of blocks d_i, group width g_i) of each stage were calculated with reference to [13] with the initial values (initial width $b_0 = 56$, slope $b_a = 38.84$, additional parameter $b_m = 2.4$, $g_0 = 16$, $d_0 = 14$). In particular, the first block at each stage used the stride-two ($s = 2$) version (Block(b)). The head network is used for predicting the acne type of the lesion patch. The fine feature maps obtained from the body network are first input to a Efficient Channel Attention Network (ECANet) [50], which avoids a dimensionality reduction and then captures the cross-channel interaction in an efficient way for improving the performance of the network. Then, global average pooling and dropout were adopted to avoid overfitting. The softmax layer was finally employed to obtain the acne classification.

The classification architecture adopted some lightweight operations, such as group convolution, which results in fewer parameters and computing costs. To further reduce app loads, network pruning was employed in the classification model. We found that the SE block in each module (block) required a mass of parameters. Thus, to accommodate these problems, ECANet was utilized to replace the SE block, which removed some less important weights from the model to generate a sparse weight matrix. Specifically, given the input feature y, the weights of channels (ω) in ECANet can be learned as follows:

$$\omega = \sigma(\mathbf{W}y) \tag{2}$$

Where σ is a Sigmoid function, $\mathbf{W}$ denotes a $K \times K$ parameter matrix and can be written as follows:

$$\begin{bmatrix} w^{1,1} & \cdots & w^{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w^{2,2} & \cdots & w^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w^{K,K-k+1} & \cdots & w^{K,K} \end{bmatrix}$$

Where K denotes channel dimension, k indicates the kernel size of convolution, and its value is 3 in this research.

Furthermore, we used smaller input image sizes (112×112) and correspondingly adjusted the size of each batch of images. The batch size was used during the training process to fit the small size input image and obtain better results. The final classification model is called the lesion regular network (LRegNet) and is more suitable for the memory space and computing power of the mobile terminal.

Finally, knowledge distillation (KD), a representative type of model compression and acceleration [51], was adopted to further reduce the model scale. As shown in Fig. 2(b), the trained LRegNet model discussed above is used as the teacher model. The student model, named Acne-RegNet, is a lesion regular network with parameters ($b_0 = 24$, $b_a = 36.44$, $b_m = 2.49$, $g_0 = 16$, $d_0 = 14$), which reduces the number of model parameters and the computing cost. To reduce the performance gap between the teacher and student models, we transferred knowledge by matching the probability distribution in feature space.

Algorithm 1 Training algorithm for Acne-RegNet model.

Input: Training set $D^{train} = \{(x_1, y_1), \cdots, (x_{N_T}, y_{N_T})\}$; val set $D^{val} = \{(x_1, y_1), \cdots, (x_{N_V}, y_{N_V})\}$; the teacher model LRegNet and = the student model Acne-RegNet.

Output: the well-trained Acne-RegNet model.

1 Pre-trained the parameters of LRegNet(Θ)with ImageNet dataset.
2 **while** *not converge* **do**
3   **for** *each batch* τ **do**
4      $D^\tau = \{(x_1, y_1), \cdots, (x_m, y_m)\} \leftarrow$ a batch of D^{train}.
5   Extract features of D^τ and obtain predictions by LRegNet.
6   Compute mean loss $L^\tau = \frac{1}{m} \sum_i L(\hat{y}_i, y_i)$ using (4).
7   Update the parameters Θ via SGD:
       $M \leftarrow \beta M + \nabla_\Theta L^\tau$
       $\Theta \leftarrow \Theta - \eta M$
8   **end for**
9   Fine-tune network parameters Θ with D^{val}.
10 **end while**
11 Pre-trained the parame ters of Acne-RegNet(Λ)
12 **while** *not converge* **do**
13   **for** *each batch* ι **do**
14     $D^\iota = \{(x_1, y_1), \cdots, (x_n, y_n)\} \leftarrow$ a batch of D^{train}.
15   Extract features $\mathbf{v}_t$ by trained LRegNet.
16   Extract features $\mathbf{v}_s$ by Acne-RegNet.
17   Calculate the probability distributions F_t and F_s
18   Obtain the predictions by Acne-RegNet.
19   Calculate the mean loss L^ι using (5)
20   Update the parameters Λ via SGD:
       $M \leftarrow \beta M + \nabla_\Lambda L^\iota$
       $\Lambda \leftarrow \Lambda - \eta M$
21   **end for**
22   Fine-tune parameters Λ with D^{val}
23 **end while**

To train the proposed Acne-RegNet model, the development set was partitioned into a training set to learn its parameters, and a validation set to tune the hyperparameters. Data augmentation was used to improve the generalization
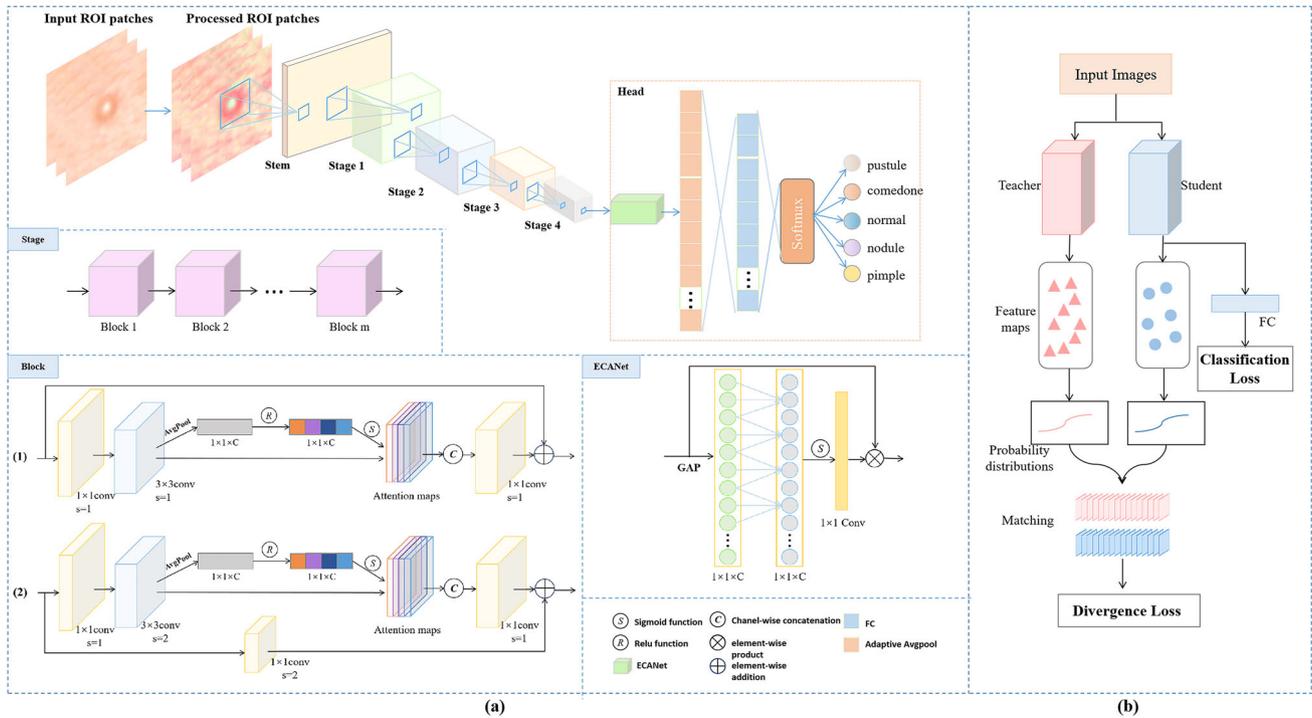
Fig. 2 Acne lesion classification architectures(a) The classification architecture consists of a stem (3 × 3 conv), gradually followed by the main body network (four stages) that performs the bulk of the computations, and then a head (ECANet followed by an adaptive average pooling and fully connected layer) that predicts five output classes. In particular, each stage is composed of a sequence of blocks that is a residual bottleneck with a group conv and SE module. (b) Take the trained network (the detailed structure shown in (a)) as the teacher model. The training images are simultaneously used as the input to the teacher model and the student model. The extracted feature maps are then transformed to probability distributions as the knowledge to supervise the training of the student model, causing the student model to obtain a comparable performance with the teacher model

and robustness (random cropping, flipping, rotating). Since the class imbalance problem was encountered during training, we adopted focal loss(FL)for alleviating the situation [52], which can be expressed as:

$$FL = -\sum_{c=1}^{C} \alpha_c y_c (1 - p_c)^{\gamma} log(p_c) \qquad (3)$$

Where C is the number of categories, $\alpha_c \in [0, 1]$ is the balance weight, y_c denotes a sign function for classification c, $p_c \in [0, 1]$ is the model's estimated probability for the class c and $\gamma \geq 0$ is tunable focusing parameter. Due to the differences in the influencing factors of the four regions (forehead: 1, nose: 2, cheek: 3, chin: 4), the loss function in this research can be expressed as follows:

$$L = \sum_{i=1}^{4} a_i I_{\{i\}} FL. \qquad (4)$$

Where a_i is the balance weight and $I_{\{i\}}$ denotes the indicator function of the facial region. In the process of knowledge distillation, the loss function consists of two losses, divergence loss and classification loss. Specifically, divergence loss is used to match the teacher model

and student model, and classification loss improves the effectiveness of the classification. The loss function L_s for training student model is shown as:

$$L_s = \lambda_1 KL(F_t, F_s) + \lambda_2 L_{ca} \qquad (5)$$

Where $KL()$ denotes the Kullback-Leibler (KL) divergence loss [53], which represents the similarity used to match the feature maps of the teacher and student models. F_t and F_s are the probability distributions of the feature space in the teacher and student models, respectively. L_{ca} is the classification loss of the student model and is expressed in (3). λ_1 and λ_2 are the balance weights. The entire training procedure for the Acne-RegNet model is summarized in Algorithm 1, where M denotes the momentum, η is the learning rate and β is the momentum constant.

3.3 App development

The proposed acne assessment method is then used to design a mobile app for automatic facial acne diagnosis. With the rapid development of the mobile Internet and Internet of Things, there is an increasing demand for the direct deployment of deep learning algorithms on smart

devices. The intelligent diagnosis of facial acne on smart devices not only is very convenient to acne patients but also reduces the pressure on dermatologists. In recent years, advances in the hardware technologies of mobile phones (such as Mate-10 of Huawei and iPhone-X of Apple) have made it possible for auxiliary diagnoses with apps. Existing mobile devices usually can only deploy a lightweight network. The proposed acne assessment method consists of two potential networks, which makes it difficult to design an app. Since the segmented areas (chin, cheek, forehead, nose) were known to people, we removed the segmentation network in the cell phone app development. Then, the acne lesion images were selected and labeled manually by users. That is, users need to circle the acne lesion by regions and upload all the lesion images. This not only reduces the model scale but it also reduces the size of the uploaded lesion images at each increment and helps save memory space, thereby meeting the need of mobile devices.

After using face photos for severity classification, users can also obtain personalized diagnosis results by filling in the metadata information, which is mainly collected in the form of multiple-choice questions and filled-in numbers. Then, a simple model utilizing multielement non-linear regression to relate the metadata to the fine-tuning score for each patient is fitted. On this basis, a mobile application for effective facial acne diagnosis has been designed and implemented.

The interface and operating instructions of the proposed app are shown in Fig. 3(a). An acne patient firstly download and install the application in their smart phones, register and log in with their QQ or WeChat. After logging into the app, the main layout, including the diagnosis and diagnosis history (including patient metadata), is shown in the treatment center. The patient can first choose the button "shooting diagnosis" and then activate the "Start Testing" button to take his/her facial photograph. After that, the
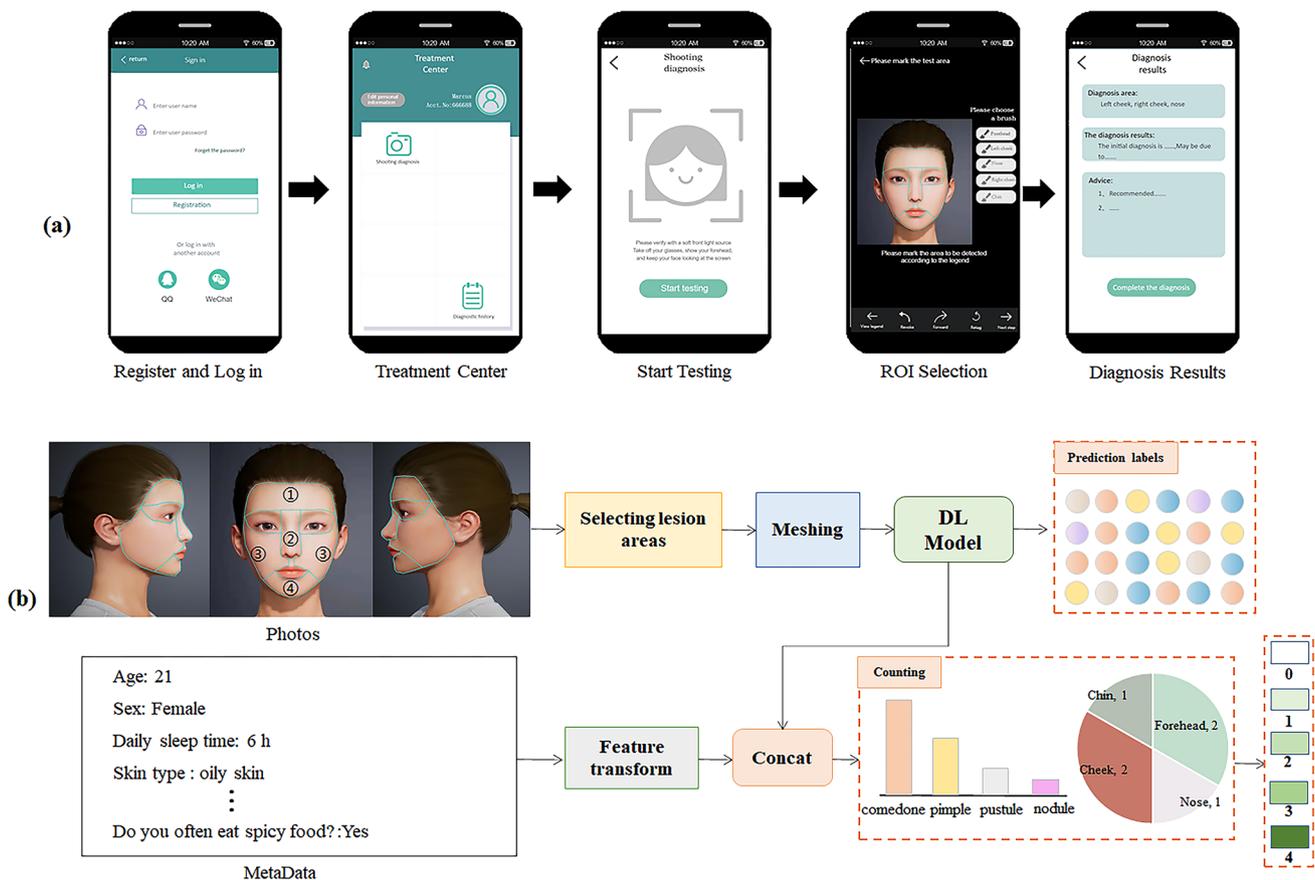

Fig. 3 Implementation of a cell phone app for facial acne grading a): An acne patient simply downloads the app onto his/her smartphone, opens the app and registers and logs into the treatment center. They can first choose the button 'shooting diagnosis', activate the 'Start Testing' button to take his/her facial photograph, and then tap on the screen to select the lesion regions. The initial diagnosis results and suggested medication can be obtained without the need for additional smartphone attachments. If they want a more personalized diagnosis, their personal metadata (see in (b)) needs to be provided in the treatment center, and then fine-tuned diagnosis results can be obtained. **b)**: After the facial photo is obtained, the user selects the lesion areas and uploads them. The uploaded images are cut to 80×80 for the input to the DL model, and the classification labels of all the patches can be obtained. Then, the severity score is calculated, and the acne grade (0,1,2,3,4) is output. If the metadata were provided, they can be transformed to fine-tune the severity score and diagnosis results

patient circles the lesion areas in each location (chin, left cheek, right cheek, forehead, nose) as the lesion images and upload them. Particularly, the uploaded images cannot include the non-skin parts, such as hair, which may cause a poor diagnosis. Finally, the app will show the initial diagnosis result, which contains the diagnosis areas, acne grading and treatment recommendations. If he/she wants a more personalized diagnosis, his/her personal metadata (see Table 2) should be provided in the treatment center (Diagnosis History), and the fine-tuned diagnosis results can be shown after several seconds. The diagnosis pipeline is illustrated in Fig. 3(**b**). The uploaded lesion image was evenly cut to 80 × 80 lesion patches for the input of the proposed Acne-RegNet model, and the classification labels of all the patches were obtained. Then, the scores of normal, comedones, pimples, pustules and nodules in each subregion were multiplied by their corresponding lesion counts and summed. The computed value is then multiplied by the region factor to obtain the severity score. If the metadata are provided, they can be transformed to fine-tune the severity score. According to the accepted medical criteria, the acne grading results can be obtained.

As shown in Fig. 4, a user can upload the facial photos after activating the 'Start testing button. Users need to mark and upload the facial images containing lesions areas. For example, the acne lesions of the patient in Fig. 4 are on both sides of the face, meaning the user needs to take two profile photos and mark the lesions areas as regions of interest (ROIs). The diagnosis results can be obtained after several minutes, which show that (1) the diagnosis areas are the left cheek, right cheek, and chin; (2) there are some comedones, 29 papules, and 3 pustules in selected ROIs, and the initial severity score and fine-tuned score are 124 and 128, respectively. The severity rating ('Severe') and advice are output according to severity criteria and treatment strategy of Table 1. Hence, the users who are concerned about facial acne can download our app for timely diagnosis and treatment referring to the obtained recommended medications. Moreover, instead of uploading the entire face image of the user, only the selected areas by user are uploaded for protecting the private information of user and accelerating the diagnosis. The app can also assist general practitioners in diagnosing facial acne, which greatly lightens the burden of dermatologists.

4 Experiment

4.1 Dataset

The facial images used in this research were collected from the ACNE04 dataset[6] and a private dataset from Xiangya Hospital in China. The private dataset characteristics are presented in Table 2. According to the ethical standards of the Medical Ethics Committee of Xiangya Hospital, all study participants provided informed consent, and all procedures were carried out in accordance with the approved guidelines and regulations.

The dataset consists of two parts, a development set for the app design and an assessment set for the clinical assessment of the app. The development set contains 1515 photos from the private dataset (January to May 2020), which was taken with an Apple iPhone 5s (Apple, Cupertino, CA), and 1455 images from the ACNE04 dataset. All the images were first randomly split into a training set, a validation set and a test set with a ratio of 8:1:1 for the fitting of the facial segmentation model. The acquired facial skin images were then cut to 80 × 80 lesion patch images. After removing the patches containing hair, the lesion patch dataset consisted of five categories: 2608 normal images, 4521 pimple images, 1290 nodule images, 1032 pustule images and 1253 comedone images
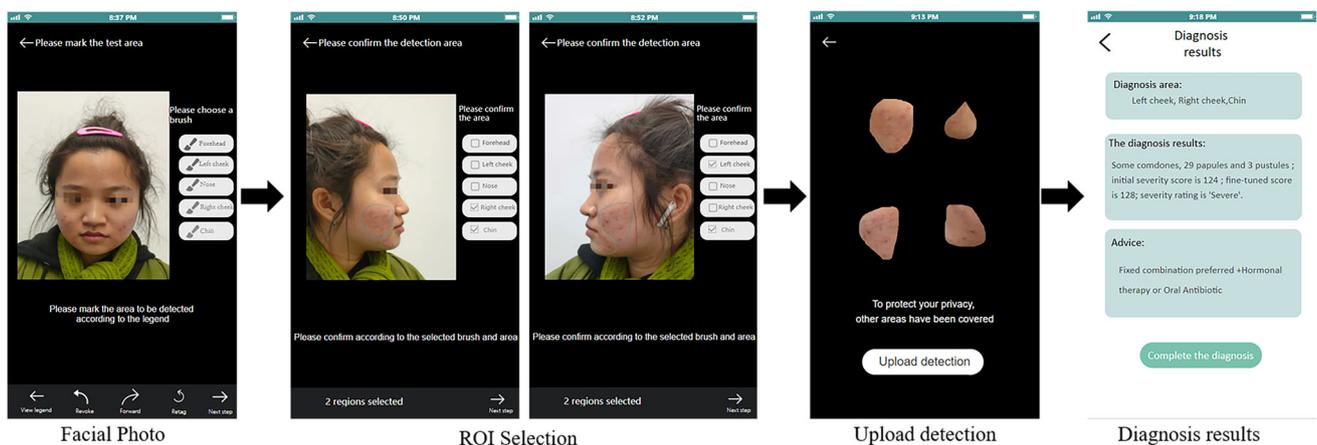

Fig. 4 Example of automatic facial acne diagnosis using app

Table 2 Private dataset characteristics

Characteristics	Development set	Assessment set
Months	(From Jan to May, 2020)	(From Jun to Aug,2020)
Number	1515	1764
Age(years):median (25th, 75th percentiles)	24(20,33)	20(19,23)
Female(%)	1014(66.93%)	1020(57.82%)
Chinese Han nationality(%)	1494(98.61%)	1728(97.96%)
BMI(m):median (25th, 75th percentiles)	20.24(18.94,22.03)	20.35(19.02,22.11)
x Skin types:		
Oily skin(%)	783(51.68%)	828(46.94%)
Dry skin(%)	123(8.12%)	264(14.96%)
Mixed skin(%)	609(40.20%)	672(38.1%)
Skin conditions based on diagnosis:		
Mild(%)	300(19.80%)	660(37.41%)
Moderate(%)	459(30.30%)	684(38.78%)
Severe(%)	573(37.82%)	360(20.41%)
Very severe(%)	183(12.08%)	60(3.40%)

BMI: Body Mass Index; Jan: January; Jun:June; Aug:august

and was randomly split into three independent subsets with a ratio of 8:1:1 for Acne-RegNet. All the lesion images were annotated by three experienced dermatologists and reviewed by a board-certified senior dermatologist with more than 25 years of experience. The assessment set contains 1764 images taken from 147 patients in different environments (camera flash on or off, indoor or outdoor, front or rear camera), which were collected from June to August 2020. Each case in the assessment set also included the corresponding metadata such as the patient demographic information (see Table 5), which were available to all clinicians and the app. This time-splitting strategy helps to simulate a study in which the model is developed based on past data and verified in future cases and is arguably a form of external verification.

All phone pictures were taken by physician assistants using all the default imaging settings. Prior to imaging, the auto-focus and brightness adjustment of the phone camera was activated and focused on the patient face by tapping the screen. Each image was taken at a distance of approximately 0.3 m from the patient face to obtain consistent images.

4.2 Evaluation metrics

To evaluate the prediction performance of the DL model, the error between the real classes and the predicted results are used on the basis of the following metrics:

(1) Mean intersection over union ($mIoU$):

$$mIoU = \frac{1}{C+1} \sum_{i=0}^{C} \frac{o_{ii}}{\sum_{j=0}^{C} o_{ij} + \sum_{j=0}^{C} o_{ji} - o_{ii}} \quad (6)$$

(2) Mean Pixel Accuracy (mPA):

$$mPA = \frac{1}{C+1} \sum_{i=0}^{C} \frac{o_{ii}}{\sum_{j=0}^{C} o_{ij}} \quad (7)$$

(3) Precision (Pre):

$$Pre = \frac{tp}{tp + fp} \quad (8)$$

(4) Recall (Rec):

$$Rec = \frac{tp}{tp + fn} \quad (9)$$

(5) F1-Score ($F1$):

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \quad (10)$$

(6) Accuracy (Acc):

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} \quad (11)$$

Where C denotes the classification of lesion patch, o_{ij} represents the number of predicted labels of j in the real label i. tp, tn, fp and fn are the number of true positives,true negatives, false positives and false negatives on lesion patchs.

4.3 Statistical analysis

To compute the confidence intervals, we used a nonparametric bootstrap procedure (1.96× standard error), with each run performed on the entire assessment set. To compare the performance of the app with the clinician, a standard one-tailed permutation test was used. Briefly, in each of the

trials, the grading result of the app was randomly swapped with itself or a comparator of the clinical diagnosis for each case, yielding an app-human difference in the accuracy sampled from the null distribution. The Kruskal-Wallis test was also adopted to analyze the difference between the assessment results. To perform the noninferiority test, the empirical P value was obtained by adding the 5% margin to the observed difference and comparing this number to its empirical quantiles.

4.4 Implementation details

The segmentation model was pretrained on the VOC dataset with the MobileNet V3 backbone [29, 49]. All the input images were first cropped to $473 \times 473 \times 3$ pixels. The adaptive moment estimation (Adam) optimizer with a mini-batch of 30 was selected, and the weight decay was set to $5e - 5$. The classification model was first pretrained on the ImageNet dataset. Before training the network, we resized the input image to $224 \times 224 \times 3$ or $112 \times 112 \times 3$ pixels and normalized it to the range of [0,1] in the RGB channels. The model was trained through a stochastic gradient descent (SGD) optimizer with a batch of 70. The momentum and weight decay were set to 0.9 and $5e - 4$, respectively. We trained the two models for 120 epochs, ensuring that the average loss on the training set was stable. The learning rate started at 0.05 and was decayed by StepLR and linear cosine, respectively. Our method ran on the NVIDIA GeForce RTX 2070 Super with 8 GB RAM, and was implemented based on the PyTorch framework.

5 Results and discussion

5.1 Model performance

To verify the effectiveness of the proposed method, the test set in the development dataset was adopted for evaluation of the two models. The experimental results are then discussed in detail.

The segmentation model achieved an mIoU of 0.9026, an mPA of 0.9336 and a parameter amount of $2.37M$, which indicated that it had far fewer trainable parameters and generated an outstanding performance in the facial region division. In particular, the lower computational overhead makes it possible to deploy the model on the terminal. The results of each subregion are shown in Fig. 5**a**. It also demonstrated the validity of the segmentation model.

The core model in our cell phone app is the Acne-RegNet model. We showed the effectiveness of each operation of

the classification model in an incremental manner. The ablation study of the refinements were compared in a test set (Table 3). The floating point operations (Flops) were used to represent the amount of computation.

First, we follow the original design of the Regular network to build our baseline, whose structure is simple and suitable for the classification task in this research. $A \rightarrow H$ operations aim to make the prediction model more efficient and effective. The first operation with a positive effect on the original model is the data preprocessing. The experimental photos suffer from noise and contrast differences, which have a negative impact on the model prediction. The median filter and contrast-limited adaptive histogram equalization (CLAHE) were employed to alleviate these problems and improve the image quality. It boosts the performance from 88.56% Acc to 91.01% Acc. The channel attention mechanism ECANet was employed to offer the potential to greatly improve the performance of the DL model. The detailed structure of ECANet is shown in Fig. 2. Unlike other attention mechanisms, it only involves a handful of parameters while providing a clear performance gain. The performance improves by 1.51%, while the parameters and Flops increase only slightly. Using the pretrained model is a common approach that can improve the feature extraction of the model. The accuracy of the classification model was further improved by 1.31%. Due to the class imbalance encountered during training, the cross entropy loss was replaced by a focal loss with a gamma of 1.5 and alpha of 0.4. On this basis, the training loss takes regions into account. Modifying the loss generally only has an impact on the training process and will not or will rarely affect the inference time. The accuracy of the model further boosts to 95.04%. However, the computational overhead also needs to be decreased to step closer to our target. The network pruning in this research used the sparse matrix to reduce the parameters and model complexity. The parameters decreased to 0.83 M, and the accuracy was only reduced by 0.18%. A larger input size may enlarge the area of the objects and lead to better prediction results but it requires more memory and is not applicable for cell phone apps. We choose a smaller input size of 112×112 to save memory. Although the accuracy decreased slightly, the Flops and inference times were reduced to 0.21G and 15 ms, respectively. This result indicated that the computational overhead is much lower. Finally, the feature-based knowledge distillation method was adopted for model compression. The detailed structure is illustrated in Fig. 2. It can obtain a simple student model whose performance is equivalent to the teacher model. The method can also reduce a large amount of the computation
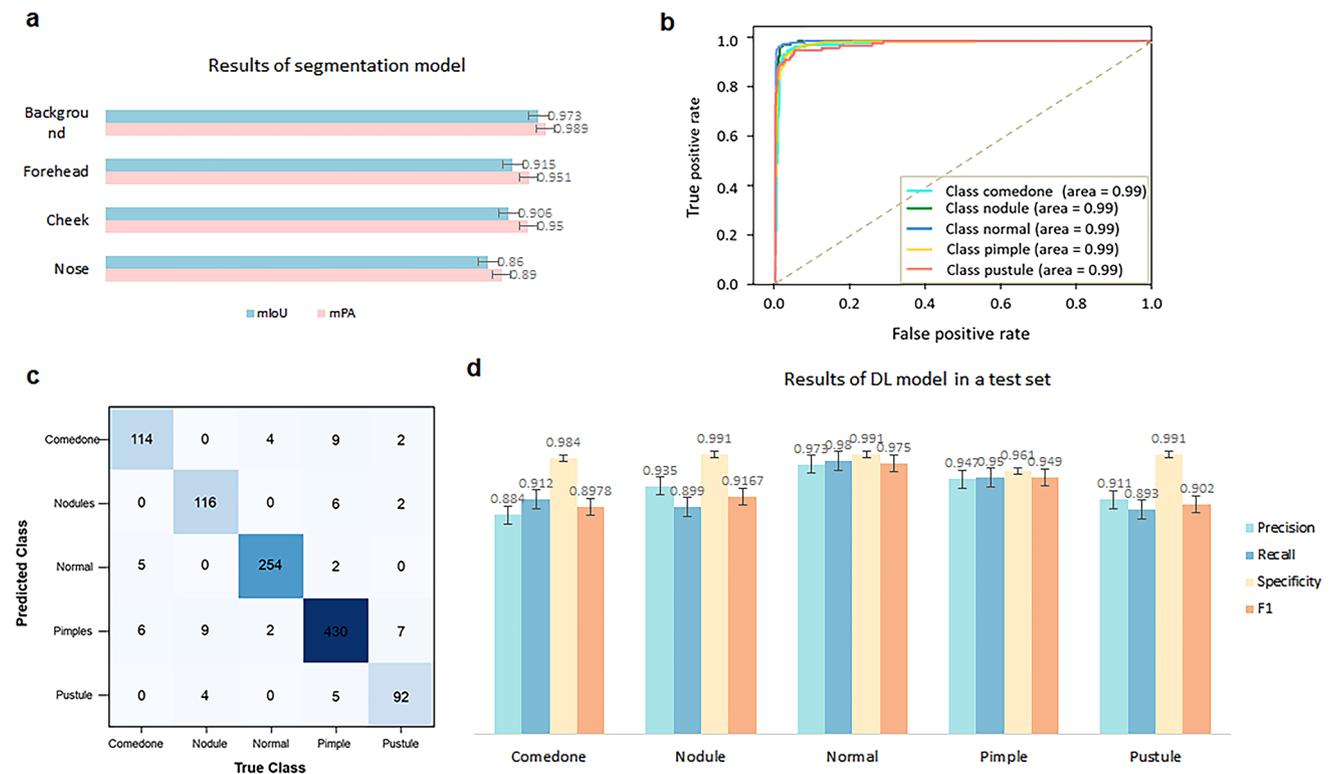
Fig. 5 Performance of segmentation model and Acne-RegNet model a. The mIoU and mPA for each region on the test set (n = 300) are presented, which range from 0 to 1. The blue and pink bars denote the mIoU and mPA, respectively. Error bars in a and d indicate 95% confidence intervals. **b.** The receiver operating characteristic (ROC) curves for comedone, nodules, normal, pimples and pustules are shown with different colors. **c.** The horizontal axis represents the ground-truth categories, and the vertical axis denotes the predicted labels. The correct classification numbers of acne, nodule, normal, acn, and pustule were 114, 116, 254, 430 and 92, respectively. **d.** The precision, recall, specificity and F1 score for each lesion class are presented, which are shown by bars of different colors

without losing accuracy. The final Acne-RegNet model achieved 2.31 M parameters, 0.05 G Flops, an inference time of 7 ms and an accuracy of 94.11%.

We further calculate the false positives, false negatives, true positives and true negatives for each class on the test set, which yields the confusion matrix and ROC curve for the test dataset (shown in Fig. 5(c) and (b)). The diagonal elements in the confusion matrix indicate the number of

correct classifications. The results showed that the model can achieve good prediction performance in each class. The ROC curves for the five lesions are close to the upper left corner, which indicates a higher classification accuracy. Figure 5(d)) shows the precision, recall, specificity and F1 score of each class. In aggregate, the results showed a promising performance: the precision was 93% (with a 95% confidence interval of 6.6%), the recall ranged from

Table 3 The ablation study of refinements on a test set

	Methods	Params (M)	Flops (G)	Infer (ms)	Acc (%)
A	RegNet	5.50	0.81	22	88.56
B	+ Data preprocessed	5.50	0.81	22	91.01
C	+ ECANet	5.50	0.81	22	92.52
D	+ Pretrained	5.50	0.81	22	93.83
E	+ Focal loss	5.50	0.81	22	95.04
F	+ Pruning	4.67	0.81	19	94.86
G	+ Smaller input	4.67	0.21	14	94.11
H	+ KD	2.31	0.05	7	94.11

Params: Parameters; Acc: Accuracy; KD: Knowledge distillation

0.893 to 0.977 in five classes, the specificity ranged from 0.961 to 0.991 and the F1 score values were from 0.898 to 0.975. The experimental analysis validates the efficiency of obtaining a promising model performance with a reduced computational cost and demonstrates that the model can be employed for clinical application.

5.1.1 Comparison with other lightweight models

The effectiveness of the designed Acne-RegNet model was compared to several state-of-the-art lightweight models, such as MobileNet V3 [29], SENet [32], EfficientNet B0 [33] and GhostNet [30], on the test dataset (Table 4). In the experiment, we optimize only the learning rate and weight decay. Other parameters (input size, data processing) have the same setup as the proposed model and enable fair comparisons. The proposed model achieves better performance than existing lightweight architectures with significantly less complexity. One possible reason is that these competing models have difficulty efficiently memorizing the training data due to their limited model capacity and may lose considerable amounts of feature information while reducing the parameters and computation. The method of compressing network parameters and Flops in the proposed model is by knowledge distillation, in which a lightweight model with comparable performance to a large model can be trained. In contrast to these lightweight networks, our designed network could memorize the training data effectively and achieve higher accuracy in the test set. The results demonstrated that the model is suitable for acne lesion classification in both accuracy and computational cost.

5.1.2 Robustness of the Acne-RegNet model

We further assess the robustness of the proposed methods to the reduction of training data by changing f in the f-fold cross-validation. The number of training samples in each fold reduces from 80% to 75%, 70%, 65% and 60% of the total lesion dataset, and f is changed from 5 to 3. As shown in Table 7, the performances of the models drop with the reduction in the training data. However, the proposed

Acne-Regnet model can still attain better performance than existing models and obtain promising results even when the training data are reduced to 60%. It is noteworthy that the MobileNet and SENet models do not perform as well as the other models, probably because they experience overfitting because of the limited samples. The experimental results showed that the proposed method is robust and feasible.

5.2 Importance of input data

We examined the importance of different input data to the cell phone app in acne grading. The 29 types of metadata and facial images for 147 acne patients were utilized for an assessment, and the comparison of the grading results is shown in Fig. 6. As discussed above, the DL model achieves a noticeable performance in ROI patch prediction; thus, the smartphone application can attain a high accuracy when only inputting images, with a value of 0.9319. The performance of the app improved slightly when all the metadata information was provided. Since there may be weakly related or interactively related factors in the metadata, we perform variable selections of this information. Random forest can be used to estimate the importance of variables in the prediction task by directly measuring the impact of each feature on the accuracy. Its basic idea is to rearrange the order of a certain column of feature values and observe how much the accuracy of the model is reduced. We then selected clinical metadata with a score greater than 5 based on the results, and the selected variables are shown in Fig. 6(a). The performance of the app improved dramatically with a value of 0.9456 when using the selected metadata.

5.3 App performance compared with clinicians

To validate the practical application value of the proposed mobile app, we compare the app performance to clinicians using the assessment set. Twelve board-certified clinicians of four different levels were included: Three professional dermatologists with more than 15 years of experience, three middle care physicians (MCPs) with more than 5

Table 4 Comparison with state-of-the-art lightweight models

Methods	Params(M)	Flops(G)	Acc(training set)	Acc(test set)
MobileNet-V3 [29]	4.21	0.22	92.23%	88.11%
SENet [32]	**1.85**	0.25	90.94%	84.93%
EfficientNet-B0 [33]	4.01	0.81	94.61%	92.32%
GhostNet [30]	2.33	0.18	94.22%	92.14%
Acne-RegNet	2.31	**0.05**	**96.75%**	**94.11%**

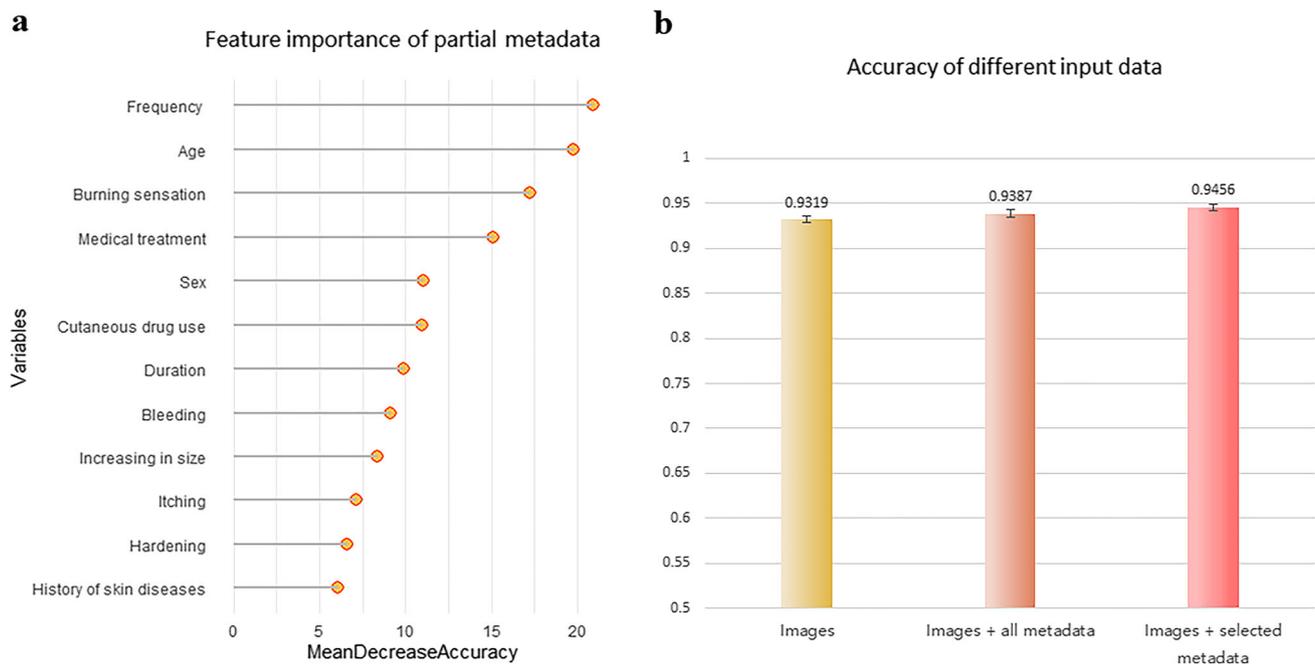Params: Parameters; Acc: Accuracy

a

b

Fig. 6 Importance of different inputs to the proposed frame

years of experience in acne diagnosis, three primary care physicians (PCPs) and three nurse practitioners (NPs). They independently graded the 147 patients using the face images and the corresponding metadata. The rating result for each patient was taken as the mode of grading by the three clinicians. We show the grading results of the app and clinicians in Fig. 7. It can be observed that clinical experience is of great importance in the acne severity assessment procedure. The app achieved an accuracy of 0.9456 (95% CI: 0.8956, 0.9762) compared to 0.9796 (95% CI: 0.9415, 0.9958) for dermatologists, 0.9048 (95% CI: 0.8454, 0.9469) for MCPs, 0.7823 (95% CI: 0.7068, 0.8461) for PCPs and 0.6531 (95% CI: 0.5702, 0.7296) for NPs. The app was obviously noninferior to the 3 MCPs. In addition, the accuracies for each dermatologist were 0.9252 (95% CI: 8701, 0.9621), 0.8916 (95% CI: 0.8495, 0.9331) and 0.9048 (95% CI: 0.8617, 0.9469). The gap in their results was due to differences in diagnostic experience and individual subjectivity. The Kruskal-Wallis statistic for the grading results of the three dermatologists and the cell phone app achieved a value of 2.7943 (P=0.0946), which indicated that the acne app was noninferior to each dermatologist. The results demonstrated that our method achieves dermatologist-level performance and even exceeds each dermatologist to a certain degree. The precision, recall, specificity and F1 score metrics were also adopted for comparison. As shown in Fig. 7, the app can achieve a performance comparable to that of dermatologists. Moreover, the values of all the metrics were superior for

MCPs, PCPs and NPs. While acne severity assessment is still a challenging task, lesion counting in each location of the face can also provide a confident reference for facial acne grading. The experimental analysis demonstrated the effectiveness of the proposed app in clinical application (Table 5).

5.4 Comparison with other studies

We compared the performance of our method with those of other existing studies in acne severity grading, as shown in Table 6. The grading criteria used in these studies varied. The research based on the Investigator's Global Assessment (IGA) criterion used DL models (Inception v4, ResNet18) to assess facial acne severity and trained the networks with the entire face skin image as input to the network [9]. Since the training samples were limited, this research did not achieve a high accuracy. Yang et al. first divided each image of acne vulgaris into four regions (chin, cheek, forehead and nose) and then combined them to form a complete facial region as the input to a convolutional neural network (CNN) [11]. All the acne images were rated in accordance with the Chinese guidelines, and the CNN model (Inception-v3) finally outputted the grading results. Nguyen et al. graded facial acne by detecting lesions and counting their number [54]. Compared with these studies, the proposed method has more complexity in conducting ratings and is more similar to the dermatologists?diagnoses. Overall, our method is of great value for the auxiliary diagnosis of facial acne.
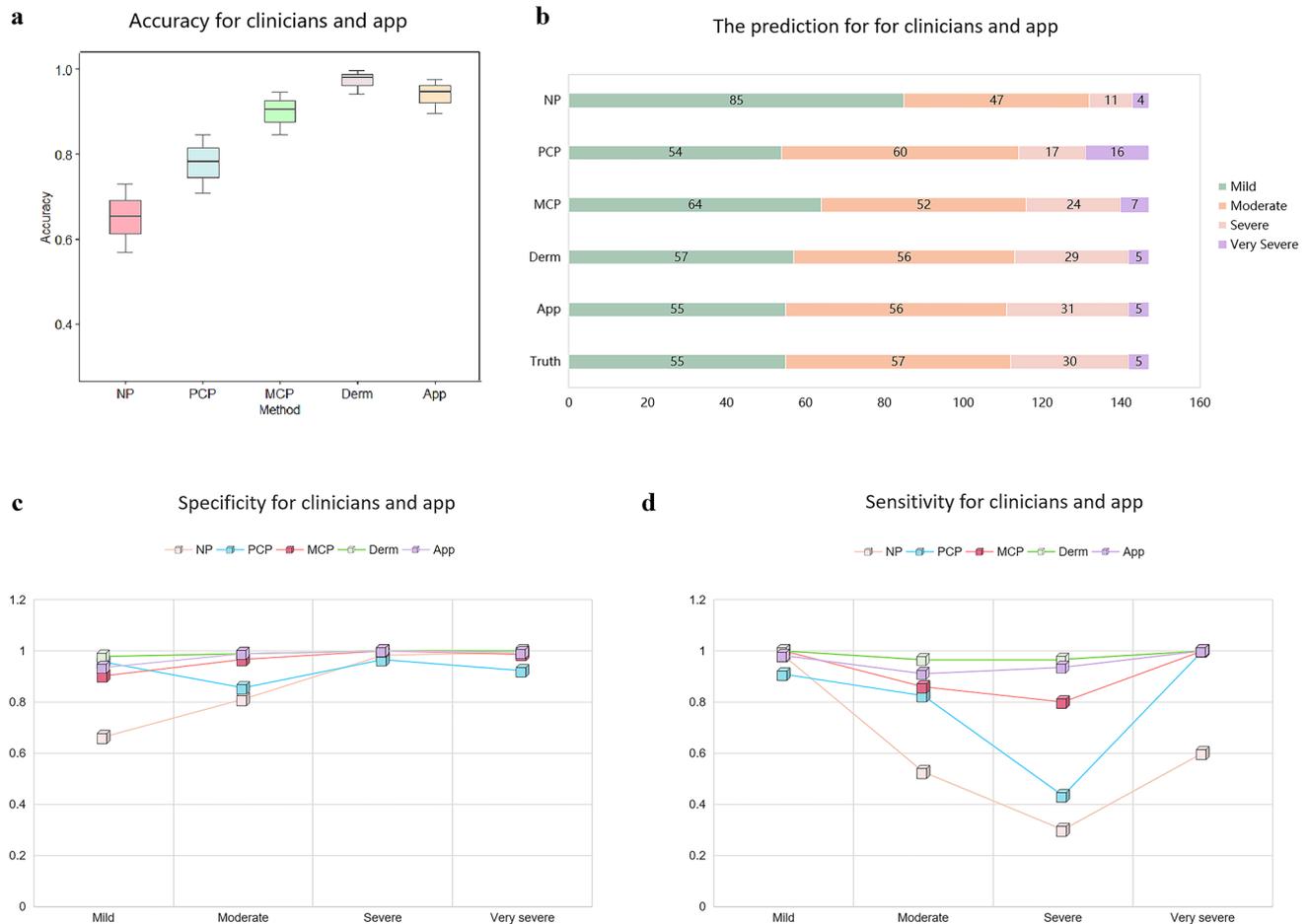
Fig. 7 Comparison with dermatologists (Derms), primary care physicians (PCPs), middle care physicians (MCPs) and nurse practitioners (NPs)

5.5 Influence of interference sources

Since shooting photos could make a difference to the app diagnosis, we compared the diagnosis of 147 patients with front and back phone photos. As shown in Fig. 8(a) and Fig. 8(b), the performance of the front-facing camera was significantly worse than that of the rear-facing camera for both the acne app diagnosis and the clinician diagnosis. The pixels of front-facing camera phones are often lower, and the area of the acne lesions is small, making accurate diagnoses difficult. Moreover, the front-facing camera of

some phones can lessen facial blemishes because of the camera beautification effect, which increases the difficulties in diagnosis. Thus, we suggest that users take facial photos with a rear camera.

In addition, the acne app should work under various background lighting conditions to be useful in a dynamic clinical environment. To that end, we collected patient photos with different backgrounds: indoors with an incandescent lamp on, outdoors and under lower light conditions with a flash on. As shown in Fig. 8(a), their accuracies were close to each other and achieved values of

Table 5 Performance robustness of models when training data reduced from 80% to 60%

Methods	Accuracy				
	0.80(5-fold,%)	0.75(5-fold,%)	0.70(3-fold,%)	0.65(3-fold,%)	0.60(3-fold,%)
MobileNet-V3 [29]	89.01±0.94	87.63±1.10	87.24±0.90	86.78±0.65	86.10±0.78
SENet [32]	84.89±1.14	83.93±0.99	83.17±0.76	82.42±0.83	82.34±0.83
EfficientNet-B0 [33]	92.21±0.90	91.34±1.16	91.14±0.92	90.96±0.88	90.74±0.97
GhostNet [30]	92.31±1.10	91.51±1.11	91.25±0.94	90.79±0.91	90.30±0.90
Acne-RegNet	94.13±0.92	93.61±1.04	93.14±0.98	92.80±0.79	92.39±0.80

Table 6 Comparison of methods used for acne severity assessment

Reference	Method	Data	Criterion	Accuracy
Lim et al. [9]	(Inception v4,ResNet18)	472 phone images	IGA	(67%,64%)
Yang et al. [11]	Inception-v3	1957 camera images	Chinese guidelines	80.00%
Nguyen et al. [54]	Multi-task Learning	1457 phone images	Hayashi criterion	83.83%
Our method	Acne-RegNet	4734 phone images	proposed criterion	93.19%
Our method	Acne-RegNet	4734 phone images		
+ Metadata	proposed criterion	94.56%		

IGA:Investigator's Global Assessment

0.932 (95% CI: 0.8785, 0.9669), 0.9388 (95% CI: 0.887, 0.9716), and 0.9456 (95% CI: 0.8956, 0.9762) in the three lighting conditions. The Kruskal-Wallis test statistic value was 0.078106 (P =0.9617), which indicated the diagnosis of the mobile phone app had no significant difference in these lighting conditions. Furthermore, Fig. 8(b) illustrates the comparison with the clinicians in different backgrounds. The results showed that the app could also attain a comparable performance with the dermatologists and exceed the MCPs, PCPs and NPs in each light condition. This further demonstrated the clinical value of the proposed cell phone app.

The skin tone of the subjects varied, so the contrast of the photos was also different. To address this issue experimentally, the contrast limited adaptive histogram equalization (CLAHE) method was adopted for adjusting the image contrast, and then images were converted into
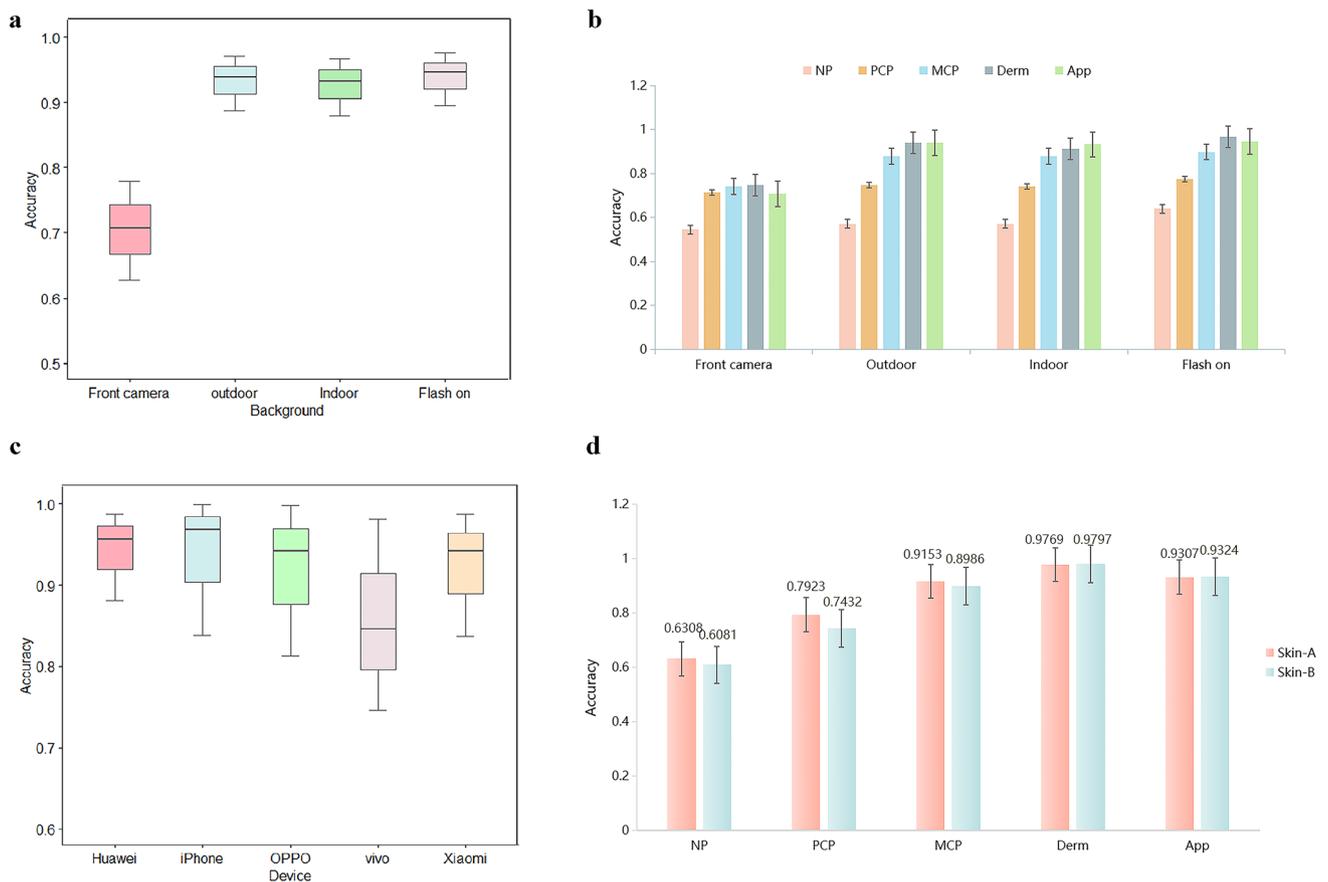

Fig. 8 Influence of interference sources: performance with different lighting conditions (a, b), different devices (c) and different skin tones (d)

the YUV color space, a commonly used color quantification system. From the results of the ablation study in Table 3, this data processing method improved the performance of the app. Theoretically, it also renders the acne app insensitive to the subject skin tone to a certain degree. The experimental results in Fig. 8(d) verified this perspective. We chose the subjects with darker complexions and fair complexions for a contrastive analysis and labeled them as A and B, respectively. The Kruskal-Wallis statistic for comparing the diagnosis of A and B in the cell phone app achieved 0.0142 (P = 0.9051). The results indicated that subject skin tone has little impact on the ability of the mobile phone system to assess the acne, which was consistent with the clinician results, as shown in Fig. 8(d).

One limitation in this study was derived from the use of a single mobile phone model and the testing of the acne grading algorithm. For wide applicability of the acne app, the algorithm should function in many types of mobile phones. To this end, we also tested the efficacy of the grading algorithm in which 30 acne patients used the app for self-diagnosis using multiple models and manufacturers of cell phones. Several prevalent mobile brands in China were chosen for this experiment: Huawei, iPhone, Oppo, Vivo, and Xiaomi. The accuracy of their diagnoses is shown in Fig. 8(c). The Kruskal-Wallis test with a value of 1.1817 (P=0.1486) indicated that there was no significant difference between them. This contrastive analysis makes it possible for widespread use of the acne app. Through the mobile phone app, patients can conduct a self-diagnosis anytime and anywhere, which will greatly reduce the diagnosis burden for the doctors and save patients the trouble of going to the hospital for treatment.

6 Conclusions

In this work, we introduced a comprehensive acne severity criterion and designed a deep learning based mobile phone app for automatic acne severity assessments. Oriented by the proposed medical criterion, our method first cropped the lesion patches at each location, and then the five classifications of patches with CNN-based lightweight architectures were investigated. The regular network was used as the baseline, and the Acne-RegNet model was

built by integrating a median filter and contrast-limited adaptive histogram equalization to improve the image quality, a channel attention module to improve the model representational power, a region-based focal loss to alleviate the problem of class imbalance and model pruning and feature-based knowledge distillation to reduce the model parameters. The effectiveness of the proposed model was analyzed with several metrics, and it improved the accuracy of acne lesion classification when compared with the state-of-the-art lightweight models. Combining the metadata of patients, a personalized diagnosis was achieved, and a mobile phone app was designed and implemented with the proposed acne grading method. We verified the effectiveness of the app by conducting various comparative experiments. It was found from the experiments that the app could attain a proximity level similar to the dermatologists and was more accurate than general practitioners. Moreover, it has a low sensitivity to background lighting conditions and skin tone in that it can maintain a stable model performance in many situations, thus it is highly practical. This research could help improve the accuracy of non-dermatologists for patient cases and represents a deep exploration of automatic diagnoses for skin diseases.

This work brings us one step closer to being able to automatically diagnose acne with a cell phone app but it also has several limitations. One limitation is that the dataset was mainly collected from one hospital, and the subjects were mainly Chinese people, which has certain regional limitations. In addition, there is a controversial issue that our grading system does not take into account the scars caused by acne. In actual clinical diagnoses, some dermatologists who perform plastic surgery consider the scars caused by acne a main indicator of the severity of acne. Another limitation is that users need to circle the acne lesions by regions and upload all the lesion images since it is difficult for the mobile terminals to deploy two DL models, which affects the simplicity of the app. Future work should focus on the scars caused by acne. We will also, in the future, combine the segmentation and classification networks for computer-side diagnoses based on the app.

Appendix A

Table 7 Metadata information

Name	Description	Possible values
Patient demographics		
Age	The age of the patient.	A float value ranging from 10 to 60.
Sex	The sex of the patient.	One of: [Female | Male]
Ethnicity	The ethnicity of the patient.	56 ethnic groups in China
History of the acne		

Table 7 (continued)

Name	Description	Possible values
Symptoms	The acne symptoms perceived by the patient.	A list of 5 symptoms (hardening , bleeding, increasing in size, itching, burning) with each symptom being one of: [Yes | No | Unknown].
Duration	The time that the acne has persisted.	One of: [Two weeks | Two to four weeks | One month | One to three months | Three to six months | More than six months | Unknown]
Frequency	Frequency of occurrence of the skin disease.	One of: [Always | Often | Occasionally | Unknown]
History of skin diseases		
Personal history	Personal past history of skin diseases.	A list of four aspects of the personal history (skin cancer, melanoma, eczema, psoriasis) with each being one of [Yes | No | Unknown].
Family history	Family past history of skin diseases.	A list of four aspects of the family history (skin cancer, melanoma, eczema, psoriasis) with each being one of [Yes | No | Unknown].
Patient state		
Allergy	Medications the patient is allergic to.	A list of 4 allergies (penicillin, cephalosporin, sulfa, nitroimidazole) with each being one of [Yes | No | Unknown].
Drug	If the patient is currently taking any medications.	One of [Yes | No | Unknown]
Pregnancy	If the patient is pregnant.	One of [Yes | No | Unknown]
Medical problem	If the patient currently has any medical problems.	One of [Yes | No | Unknown]
Previous treatment state		
Biopsy	If there has been a previous biopsy	One of [Yes | No | Unknown]
Past medication	If the patient used medications for the skin problem.	One of [Yes | No | Unknown].
Condition after treatments	Progression of the skin problem If the patient received treatment before.	One of: [Improved | Worsened | Unknown]

Declarations

Conflict of Interests The authors declare no competing interests.

References

1. Yim M-J, Lee JM, Kim H-S, Choi G, Kim Y-M, Lee D-S, Choi I-W (2020) Inhibitory effects of a sargassum miyabei yendo on cutibacterium acnes-induced skin inflammation. Nutrients 12(9):2620
2. Zaenglein AL (2018) Acne vulgaris. N Engl J Med 79(14):1343–1352
3. Oulès B, Philippeos C, Segal J, Tihy M, Rudan MV, Cujba A-M, Grange PA, Quist S, Natsuga K, Deschamps L (2020) Contribution of gata6 to homeostasis of the human upper pilosebaceous unit and acne pathogenesis. Nature Communications 11(1):1–17

4. Samuels DV, Rosenthal R, Lin R, Chaudhari S, Natsuaki MN (2020) Acne vulgaris and risk of depression and anxiety: a meta-analytic review. J Am Acad Dermatol 83(2):532–541

5. Vallerand I, Lewinson R, Parsons L, Lowerison M, Frolkis A, Kaplan G, Barnabe C, Bulloch A, Patten S (2018) Risk of depression among patients with acne in the UK: a population-based cohort study. Br J Dermatol Suppl 178(3):194–195

6. Wu X, Wen N, Liang J, Lai Y-K, She D, Cheng M-M, Yang J (2019) Joint acne image grading and counting via label distribution learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 10642–10651

7. Akpinar Kara Y, Ozdemir D (2020) Evaluation of food consumption in patients with acne vulgaris and its relationship with acne severity. J Cosmet Dermatol 19(8):2109–2113

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

9. Lim ZV, Akram F, Ngo CP, Winarto AA, Lee WQ, Liang K, Oon HH, Thng STG, Lee HK (2020) Automated grading of acne vulgaris by deep learning with convolutional neural networks. Skin Res Technol 26(2):187–192

10. Jung C, Yeo I, Jung H (2019) Classification model of facial acne using deep learning. J Korea Inst Inf Commun Eng 23(4):381–387

11. Yang Y, Guo L, Wu Q, Zhang M, Zeng R, Ding H, Zheng H, Xie J, Li Y, Ge Y et al (2021) Construction and evaluation of a deep learning model for assessing acne vulgaris using clinical images. Dermatology and Therapy, pp 1–10

12. Seite S, Moyal D, Abidi K, Le Dantec G, Khammari A, Benzaquen M, Dréno B (2020) 14034 development and accuracy of an artificial intelligence algorithm for acne evaluation. J Am Acad Dermatol 83(6):17

13. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P (2020) Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 10428–10436

14. Muthusamy D, Rakkimuthu P (2022) Steepest deep bipolar cascade correlation for finger-vein verification. Appl Intell 52(4):3825–3845

15. Mittal H, Pandey AC, Pal R, Tripathi A (2021) A new clustering method for the diagnosis of covid19 using medical images. Appl Intell 51(5):2988–3011

16. Abbas A, Abdelsamea MM, Gaber MM (2021) Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. Appl Intell 51(2):854–864

17. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, Kanada K, de Oliveira Marinho G, Gallegos J, Gabriele S (2020) A deep learning system for differential diagnosis of skin diseases. Nat Med 26(6):900–908

18. Maron RC, Haggenmüller S, Von Kalle C, Utikal JS, Meier F, Gellrich FF, Hauschild A, French LE, Schlaak M, Ghoreschi K (2021) Robustness of convolutional neural networks in recognition of pigmented skin lesions. Eur J Cancer 145:81–91

19. Dulmage B, Tegtmeyer K, Zhang MZ, Colavincenzo M, Xu S (2021) A point-of-care, real-time artificial intelligence system to support clinician diagnosis of a wide range of skin diseases. J Investig Dermatol 141(5):1230–1235

20. Wu H, Pan J, Li Z, Wen Z, Qin J (2020) Automated skin lesion segmentation via an adaptive dual attention module. IEEE Trans Med Imaging 40(1):357–370

21. Wang Z, Keane PA, Chiang M, Cheung CY, Wong TY, Ting DSW (2020) Artificial intelligence and deep learning in ophthalmology. Artif Intell Med, pp 1–34

22. Schmidt-Erfurth U, Reiter GS, Riedl S, Seeböck P, Vogl W-D, Blodi BA, Domalpally A, Fawzi A, Jia Y, Sarraf D et al (2021) Ai-based monitoring of retinal fluid in disease activity and under therapy. Prog Retin Eye Res, pp 100972

23. Phaphuangwittayakul A, Guo Y, Ying F, Dawod AY, Angkurawaranon S, Angkurawaranon C (2021) An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head ct images for traumatic brain injury. Appl Intell, pp 1–19

24. Lian J, Liu J, Zhang S, Gao K, Liu X, Zhang D, Yu Y (2021) A structure-aware relation network for thoracic diseases detection and segmentation. IEEE Trans Med Imaging 40(8):2042–2052

25. Shen B, Zhang Z, Shi X, Cao C, Zhang Z, Hu Z, Ji N, Tian J (2021) Real-time intraoperative glioma diagnosis using fluorescence imaging and deep convolutional neural networks. European Journal Of Nuclear Medicine And Molecular Imaging 48(11):3482–3492

26. Jie B, Liu M, Lian C, Shi F, Shen D (2020) Designing weighted correlation kernels in convolutional neural networks for functional connectivity based brain disease diagnosis. Medical Image Analysis 63:101709

27. Chaddad A, Hassan L, Desrosiers C (2021) Deep radiomic analysis for predicting coronavirus disease 2019 in computerized tomography and x-ray images. IEEE Trans Neural Netw Learn Syst 33(1):3–11

28. Kulkarni U, Meena S, Gurlahosur SV, Bhogar G (2021) Quantization friendly mobilenet (qf-mobilenet) architecture for vision based applications on embedded platforms. Neural Netw 136:28–39

29. Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2022) A survey of modern deep learning based object detection models. Digit Signal Process, pp 103514

30. Li Z, Liu F, Yang W, Peng S, Zhou J (2021) A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst

31. Kaddar B, Fizazi H, Hernández-Cabronero M, Sanchez V, Serra-Sagristà J (2021) Divnet: efficient convolutional neural network via multilevel hierarchical architecture design. IEEE Access 9:105892–105901

32. Hu J, Shen L, Sun G (2020) Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell, pp 42(8)

33. Dhillon A, Verma GK (2020) Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell 9(2):85–112

34. Passalis N, Tefas A (2018) Training lightweight deep convolutional neural networks using bag-of-features pooling. IEEE Trans Neural Netw Learn Syst 30(6):1705–1715

35. Paluru N, Dayal A, Jenssen HB, Sakinis T, Cenkeramaddi LR, Prakash J, Yalavarthy PK (2021) Anam-net: anamorphic depth embedding-based lightweight cnn for segmentation of anomalies in covid-19 chest ct images. IEEE Trans Neural Netw Learn Syst 32(3):932–946

36. Luo J-H, Wu J (2020) Autopruner: an end-to-end trainable filter pruning method for efficient deep model inference. Pattern Recogn 107:107461

37. Guo J, Zhang W, Ouyang W, Xu D (2020) Model compression using progressive channel pruning. IEEE Trans Circuits Syst Video Technol 31(3):1114–1124

38. Choudhary T, Mishra V, Goswami A, Sarangapani J (2020) A comprehensive survey on model compression and acceleration. Artif Intell Rev 53(7):5113–5155

39. Gong C, Chen Y, Lu Y, Li T, Hao C, Chen D (2020) Vecq: minimal loss dnn model compression with vectorized weight quantization. IEEE Trans Comput 70(5):696–710

40. Wang L, Yoon K-J (2021) Knowledge distillation and student-teacher learning for visual intelligence: a review and new outlooks. IEEE Trans Pattern Anal Mach Intell

41. Liu Y, Shu C, Wang J, Shen C (2020) Structured knowledge distillation for dense prediction. IEEE Trans Pattern Ana Mach Intell

42. Dou Q, Liu Q, Heng PA, Glocker B (2020) Unpaired multi-modal segmentation via knowledge distillation. IEEE Trans Med Imaging 39(7):2415–2425

43. Duarte-Rojo A, Bloomer PM, Rogers RJ, Hassan MA, Dunn MA, Tevar AD, Vivis SL, Bataller R, Hughes CB, Ferrando AA (2021) Introducing el-fit (exercise and liver fitness): a smartphone app to prehabilitate and monitor liver transplant candidates. Liver Transplant 27(4):502–512

44. Krishnamurti T, Davis AL, Rodriguez S, Hayani L, Bernard M, Simhan HN (2021) Use of a smartphone app to explore potential underuse of prophylactic aspirin for preeclampsia. JAMA Network Open 4(10):2130804–2130804

45. Freeman K, Dinnes J, Chuchu N, Takwoingi Y, Bayliss SE, Matin RN, Jain A, Walter FM, Williams HC, Deeks JJ (2020) Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. BMJ, vol 368

46. Goceri E (2021) Diagnosis of skin diseases in the era of deep learning and mobile technology. Comput Biol Med 134:104458

47. Verma S, Razzaque MA, Sangtongdee U, Arpnikanondt C, Tassaneetrithep B, Hossain A (2021) Digital diagnosis of hand, foot, and mouth disease using hybrid deep neural networks. IEEE Access 9:143481–143494

48. Thiboutot DM, Dréno B, Abanmi A, Alexis AF, Araviiskaia E, Cabal MIB, Bettoli V, Casintahan F, Chow S, Da Costa A (2018) Practical management of acne for clinicians: an international consensus from the global alliance to improve outcomes in acne. J Am Acad Dermatol 78(2):1–23

49. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J (2018) A survey on deep learning techniques for image and video semantic segmentation. Appl Soft Comput 70:41–65

50. Wang J, Yu J, He Z (2022) Deca: a novel multi-scale efficient channel attention module for object detection in real-life fire images. Appl Intell 52(2):1362–1375

51. Gou J, Yu B, Maybank SJ, Tao D (2021) Knowledge distillation: a survey. Int J Comput Vis 129(6):1789–1819

52. Romdhane TF, Pr MA (2020) Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. Comput Biol Med 123:103866

53. Asperti A, Trentin M (2020) Balancing reconstruction error and kullback-leibler divergence in variational autoencoders. IEEE Access 8:199440–199448

54. Nguyen A, Thai H, Le T (2021) Severity assessment of facial acne. In: International Conference on Computational Collective Intelligence, pp 599–612. Springer

Affiliations

Jiaoju Wang[1] · Yan Luo[2] · Zheng Wang[1,3] · Alphonse Houssou Hounye[1] · Cong Cao[1] · Muzhou Hou[1] · Jianglin Zhang[4,5]

✉ Jianglin Zhang
zhang.jianglin@szhospital.com

[1] School of Mathematics and Statistics, Central South University, Changsha, 410083,Hunan, China

[2] Department of dermatology of Xiangya hospital, Central South University, Changsha, 410083,Hunan, China

[3] Science and Engineering School, Hunan First Normal University, Changsha, 410083,Hunan, China

[4] Department of Dermatology of Shenzhen People's Hospital The Second Clinical Medical College of Jinan Uninversity, The First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, 518020, Guangdong, China

[5] Candidate Branch of National Clinical Research Center for Skin Diseases, Shenzhen, 518020, Guangdong, China