# Facial Acne Classification Using EfficientNetV2-S Deep Learning Model

## Abstract

Acne vulgaris is a prevalent dermatological condition among adolescents and young adults worldwide. Accurate classification of acne lesions plays an essential role in determining appropriate treatments and supporting telemedicine applications. This study proposes an automated multi-class facial acne classification model employing the EfficientNetV2-S convolutional neural network (CNN) architecture. The dataset comprises 4,673 annotated facial images categorized into five classes: blackheads, whiteheads, papules, pustules, and cysts. All images were preprocessed through resizing, normalization, and augmentation techniques to improve robustness. Transfer learning with fine-tuning was adopted to leverage pre-trained weights from ImageNet and adapt the model for dermatological features. Evaluation on a separate test set demonstrated an accuracy of 97.29%, with macro-average precision, recall, and F1-score all reaching 0.97. Compared to conventional architectures such as ResNet50 and VGG16, the proposed approach achieves higher accuracy with reduced computational complexity. These results demonstrate the effectiveness of EfficientNetV2-S for facial acne classification and its potential for implementation in clinical decision support systems and mobile health solutions targeting acne classification.

Keywords—Acne classification, Deep learning, EfficientNetV2-S, Convolutional neural networks, Transfer learning.

## I. Introduction

Skin disease detection has become a prominent research focus in medical image analysis, particularly due to its impact on public health and the potential of deep learning to improve diagnostic accuracy [1]–[3]. Acne vulgaris is one of the most prevalent dermatological conditions, affecting over 85% of adolescents and young adults [4]. While typically non-life-threatening, acne has profound psychosocial impacts, including reduced self-esteem and increased anxiety [5].  
  
Traditional acne diagnosis relies on dermatologists’ subjective visual assessment, which can be inconsistent and inaccessible in low-resource settings [6]. Advances in convolutional neural networks (CNNs) have enabled automated classification of skin lesions with accuracy surpassing traditional methods [7]–[9]. However, many existing studies have focused on binary classification of limited acne types or have employed outdated architectures such as VGG16 [10], [11]. Furthermore, datasets used in prior studies often lack variability in lighting, image quality, and skin tones, limiting generalizability [12].  
  
Recent innovations such as EfficientNetV2 architectures have demonstrated superior performance and computational efficiency in image classification tasks [13], [14]. Despite this, limited research has evaluated EfficientNetV2 for multi-class acne classification across diverse real-world images.  
  
Therefore, this study aims to develop and validate a multi-class facial acne classification model based on the EfficientNetV2-S architecture, trained on a large and heterogeneous dataset. The main contributions are: (1) demonstrating the effectiveness of EfficientNetV2-S in accurately distinguishing five acne types; (2) comparing its performance against traditional CNN architectures; and (3) providing a reproducible workflow suitable for deployment in mobile health applications.

## II. Research Method

A. Dataset and Preprocessing  
The dataset consisted of 4,673 high-resolution facial images collected from the publicly available DermNet dataset [15], categorized into five classes: blackheads, whiteheads, papules, pustules, and cysts. Images were split into training (2,834), validation (921), and test (918) sets. Preprocessing included resizing images to 224×224 pixels, normalizing pixel values to [0,1], and applying augmentation (random rotation ±15°, zoom up to 20%, horizontal flips, brightness adjustment between 0.8–1.2).  
  
B. Model Architecture  
The EfficientNetV2-S architecture was initialized with ImageNet pre-trained weights [14]. The feature extractor was followed by: GlobalAveragePooling2D, Dense(512)+BN+Dropout(0.4), Dense(128)+BN+Dropout(0.3), and Dense(5) softmax output layer. The model contained ~21 million parameters.  
  
C. Training Procedure  
Training used categorical crossentropy loss and the Adam optimizer (lr=1e-4). EarlyStopping (patience=5), ReduceLROnPlateau (factor=0.5), and ModelCheckpoint callbacks were applied. Training ran for 20 epochs with batch size=32 on an NVIDIA Tesla T4 GPU.  
  
D. Evaluation Metrics  
Performance was evaluated on the held-out test set using accuracy, precision, recall, F1-score, confusion matrix, and visual inspection.