Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Application of Weighted Loss Function in Convolutional Neural Network for Acne Image Classification

Abubakar Sidik[1*], Ade Irma Purnamasari[2], Denni Pratama[3], Puji Pramudya Marta[4], Yudhistira Arie Wijaya[5]

[1,2]Computer Science, STMIK IKMI Cirebon
[3] Information Management, STMIK IKMI Cirebon
[4,5]Information Systems, STMIK IKMI Cirebon
*abubakarsidiknevada@email.com[*1], irma2974@yahoo.com [2],*
pratamadenni@gmail.com [3],prammarta88@gmail.com[4], yudhistira010471@gmail.com [5]

Abstract

Automated acne image classification using Convolutional Neural Networks (CNN) holds significant potential in dermatological diagnosis but faces a fundamental challenge of class imbalance. This phenomenon causes standard models to be biased towards majority classes and fail to recognize clinically important minority classes. This study aims to address this bias by applying a Weighted Loss Function to the EfficientNetB1 architecture. The research method employs a comparative experimental approach between two scenarios: the Baseline model (Standard Cross-Entropy) and the Proposed model (Weighted Cross-Entropy). The dataset consists of 5 acne classes with an imbalanced distribution. The results show that the Weighted Loss model significantly outperforms the Baseline model. Overall accuracy increased from 80% to 86%. The most significant improvement occurred in the minority class 'Papules', where the F1-Score surged by 0.10 points (from 0.71 to 0.81). It is concluded that the application of Weighted Loss Function effectively overcomes bias due to imbalanced data without the need for synthetic data augmentation, resulting in a fairer and more reliable model for clinical implementation.

Keywords: *Acne Classification, Class Imbalance, Convolutional Neural Network, EfficientNetB1, Weighted Loss Function*

1. Introduction

The application of *Deep Learning*, particularly *Convolutional Neural Networks* (CNN), has brought significant advances in automated dermatological diagnosis. Although CNN excels at extracting visual features hierarchically, its performance is highly dependent on the quality of the training data [1]. The main challenge in medical image analysis today lies not only in the model architecture, but also in the characteristics of the dataset itself. Medical images often suffer from labeling subjectivity and *label noise* due to differences in clinical perception among dermatologists, which can degrade model accuracy [2], [3].

In addition to label quality issues, the most persistent technical obstacle is class *imbalance*. This phenomenon is common in medical datasets where certain pathological conditions (minority classes) are much less common than normal or mild cases (majority classes) [4]. In the context of acne (*Acne Vulgaris*) classification, data distribution is often skewed, where severe inflammatory lesions such as cysts or nodules are very limited in number compared to comedones [5].

The presence of this class imbalance has serious consequences. Standard learning algorithms that are not mitigated tend to be biased towards the majority class in order to minimize global *loss* [6]. As a result, models can achieve *artificially high* overall accuracy, but fail completely in recognizing minority classes, which are often critical conditions that require medical treatment [7].

To address this issue, the *Cost-Sensitive Learning* approach has emerged as an efficient algorithmic solution. Unlike *oversampling* techniques that increase computational load, this method modifies the objective function by assigning a greater penalty weight to minority class classification errors [8]. Previous studies have shown that *Weighted Cross-Entropy* (WCE) is effective in forcing models to learn features from underrepresented classes without the need to manipulate physical data [9], [10].

On the other hand, the choice of *backbone* architecture is also crucial. *EfficientNet* has been proven to deliver superior performance compared to its predecessor architectures (such as ResNet or VGG) thanks to its *compound scaling* method, which efficiently balances the depth, width, and resolution of the network [11]. The *EfficientNet-B1* variant, in particular, offers an optimal balance between accuracy and computational speed for medium-scale datasets [12].

Although the potential of *Weighted Loss* and *EfficientNet* has been recognized separately, research that specifically integrates the two for multi-class acne classification is still limited. This study aims to fill this gap by evaluating the effectiveness of applying *the Weighted Loss Function* to the *EfficientNetB1* architecture to improve detection sensitivity in minority acne classes.

2. Research Method

This study uses a quantitative experimental design to compare the performance of two model training scenarios on the same dataset.

2.1. Data Collection and Preparation

The dataset used is secondary acne image data consisting of five diagnostic classes: *Blackheads, Cysts, Papules, Pustules,* and *Whiteheads*. Based on preliminary analysis (*Exploratory Data Analysis*), a significant imbalance was found, with the ratio between the majority class (*Blackheads*) and the minority class (*Whiteheads*) reaching approximately 5.6:1.

The data was divided using the *Stratified Split* method with a proportion of 70% training data, 20% validation data, and 10% test data. This method ensures that the imbalanced class proportions are maintained in each data *subset*.

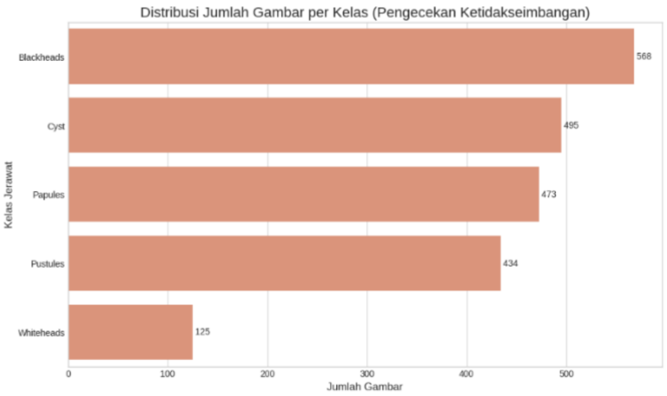

Fig. 1: Number of data sets used

2.2. Data Preprocessing

All images underwent *preprocessing* in the form of pixel value normalization (scale 1/255) and size standardization (*resizing*) to 150x150 pixels. Specifically for the training data, data augmentation (rotation, *zoom*, and *flip*) was applied to increase feature variation and prevent *overfitting.*

2.3. Model Architecture

This study adopts a *Transfer Learning* approach using EfficientNetB1 as the basic architecture (*backbone*) for feature extraction. This model has been *pre-trained* on the large-scale ImageNet dataset. EfficientNetB1 was chosen based on its advantage in applying the *compound scaling* method, which optimally balances the depth, width, and resolution of the network using combined coefficients. This approach allows the model to achieve high accuracy with much better computational efficiency compared to conventional CNN architectures.

To adapt the model to the specific task of acne classification, the classification part (*top layer*) was modified with the following *custom head* architecture:

1. Global Average Pooling (GAP): Used to reduce the spatial dimensions of *feature maps* into one-dimensional vectors, which significantly reduces the number of parameters and the risk *of overfitting*.
2. Dense Layer (Fully Connected): A hidden layer with a ReLU (*Rectified Linear Unit*) activation function is added to learn complex non-linear feature combinations from the visual representation of acne.
3. Dropout: A regularization technique applied by randomly deactivating some neurons during training to prevent the model from becoming too dependent on certain features (*co-adaptation*).
4. Output Layer: The final layer uses a Softmax activation function that produces a probability distribution to classify images into 5 target classes (*Blackheads, Cysts, Papules, Pustules, Whiteheads*).

2.4. Experimental Scenarios

To test the effectiveness of the proposed method and validate the research hypothesis, the experiment was designed using a comparative approach (*A/B testing*) that divided the training into two parallel scenarios with identical *hyperparameter* configurations, except for the loss function.

2.4.1. Scenario A: Baseline Model (Control)

This scenario functions as an experimental control (*baseline*). In this scenario, the *EfficientNetB1* model is trained using the standard Categorical Cross-Entropy loss function without weight modification.

1. Rationale: This approach represents the general conditions for training models on datasets that are assumed to have a balanced distribution.
2. Mechanism: In this scenario, each image sample contributes equally to the total *loss* value, regardless of which class the sample belongs to.
3. Implications: Given the imbalanced dataset, this scenario is predicted to produce a model that has a strong inductive bias towards the majority class (such as *Blackheads*), because the model will prioritize minimizing global error by predicting the most frequently occurring class[1].

2.4.2. Scenario B: Proposed Model (Weighted Loss)

This scenario implements the Cost-Sensitive Learning strategy through the use of Weighted Cross-Entropy. This approach aims to intervene in the gradient optimization process by assigning different "costs" or penalties for each classification error.

1. Weight Calculation: Class weights(w_j) are calculated automatically before training begins using the inverse class frequency method from the Scikit-learn library (compute_class_weight). The weight determination formula is as follows:

$$w_j = \frac{n_{samples}}{n_{classes} \times n_{samples_j}} \tag{1}$$

Where $n_{samples}$ is the total number of images, $n_{classes}$ is the number of classes (5), dan

$$n_{samples_j}\ is\ the\ number\ of\ images\ in\ the\ class - j. \tag{2}$$

2. Mechanism: With this formula, minority classes (such as *Whiteheads* or *Papules*) will get a weight value of$w_j > 1$, while majority classes will get a weight of$w_j < 1$. When the model makes a mistake on a minority class, the *loss* value will be multiplied by that large weight, resulting in a sharper gradient[2].
3. Implications: This mechanism forces the model to pay more attention to minority classes in order to reduce *loss* values, thereby improving sensitivity and *F1-Score* without the need for physical manipulation of the dataset (such as *oversampling*)[3].

3. Results And Discussion

This section presents a comparative analysis between the Baseline model (Scenario A) and the Weighted Loss model (Scenario B). The analysis was conducted in stages, starting with an evaluation of training dynamics to detect overfitting behavior, followed by a comparison of final performance metrics, and ending with an in-depth investigation of classification error patterns using a confusion matrix.

3.1. Training Dynamics Analysis (Training History)

Analysis of the training history graph provides crucial insights into how the model learns features from unbalanced data.

3.1.1. Baseline Model Behavior

The *Baseline* model training graph (Scenario A) in Figure 1 shows strong indications of *overfitting*. In the accuracy graph, although the model is able to achieve near-perfect training accuracy (~0.99), validation accuracy stagnates at around 0.90 without a significant upward trend. This disparity is further confirmed by the *Loss* graph. While *the training loss* decreases consistently (convergence), *the validation loss* remains at a very high value (ending at around 1.75). The high residual *loss* in the validation data indicates that the model tends to memorize the training data, but has a low *confidence level* when generalized to new data.

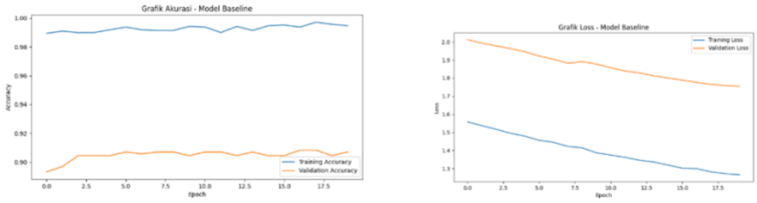

Fig. 2: Baseline Model Training History Graph (Scenario A)

3.1.2 . Weighted Loss Model Behavior

Conversely, the application of *Weighted Cross-Entropy* in Scenario B (Figure 2) shows a significant improvement in convergence stability. Although the gap between the training and validation curves still exists as a common characteristic of complex *Deep Learning* architectures, the behavior of the loss function is much more controlled. The *validation loss* value was successfully reduced to around 1.58, much lower than *the Baseline*. This decrease in *loss* magnitude proves that the weight penalty on the minority class acts as effective regularization, forcing the model to improve its prediction probability distribution and reduce inductive bias towards the majority class.

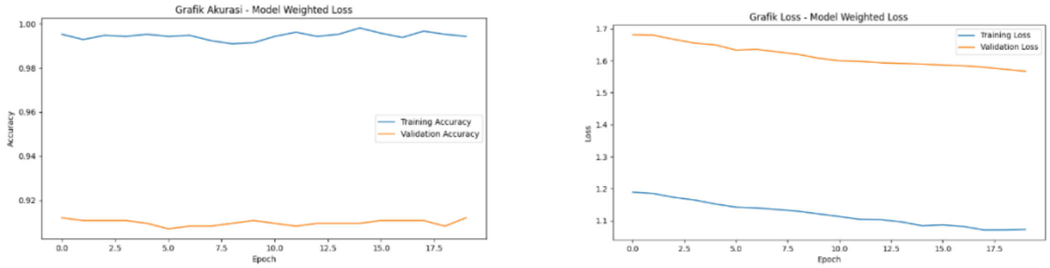

Fig. 3: Model Weighted Loss Training History Graph (Scenario B)

3.2. Scenario 1 Results: Baseline Model

The *Baseline* Model was trained using *Standard Categorical Cross-Entropy*. Based on the evaluation results on the test data, this model achieved an overall accuracy of 80%. Details of the performance per class are presented in Table 1.

Table 1. Baseline Model Performance (Without Weights)

Class	Precision	Recall	F1-Score
Blackheads	0.95	0.78	0.86
Cyst	0.89	0.74	0.81
Papules	0.73	0.70	0.71
Pustules	0.66	0.86	0.75
Whiteheads	0.67	0.97	0.79
Accuracy	—	—	**0.80**

The data in Table 1 shows performance imbalance. The *Papules* class recorded the lowest performance with *an F1-Score* of 0.71, indicating the model's difficulty in recognizing unique papule features without weight penalties. In addition, there is *an* extreme *trade-off* in the *Whiteheads* class, where *Recall* is very high (0.97) but *Precision* is low (0.67), indicating a large number of *False Positives*.
To visualize prediction errors, Figure 1 presents the *Confusion Matrix of the* Baseline model.

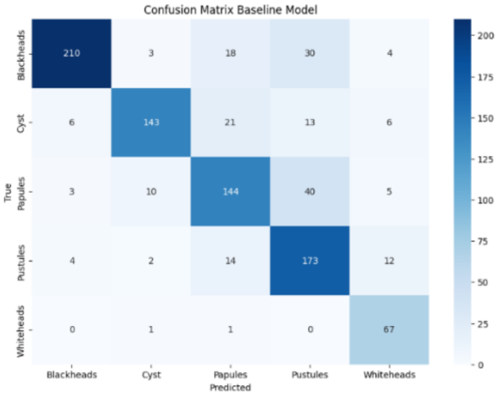

Fig. 4: Confusion Matrix of the Baseline Model

Analysis of Figure 1 shows significant visual ambiguity. It was identified that the *Baseline* model often mispredicts the *Papules* class. A total of 40 genuine *Papules* samples were misclassified as *Pustules*. This occurred due to the visual morphological similarity between inflamed papules and early-stage pustules.

3.3. Scenario 2 Results: Weighted Loss Model

The proposed model was trained by applying *Weighted Cross-Entropy*, where minority classes are given greater loss weights. The evaluation results show a significant improvement in performance, as summarized in Table 2.

Table 2. Weighted Loss Model Performance (Proposal)

Class	Precision	Recall	F1-Score
Blackheads	0.92	0.89	0.91
Cyst	0.92	0.82	0.87
Papules	0.82	0.80	0.81
Pustules	0.77	0.86	0.81
Whiteheads	0.76	0.97	0.85
Accuracy	—	—	**0.86**

The application *of Weighted Loss* successfully increased the overall accuracy to 86%. The most crucial improvement occurred in the *Papules* class, where *the F1-Score* jumped dramatically from 0.71 to **0.81**. This proves that *loss* weighting effectively forces the model to learn distinctive features in difficult classes.

The impact of this improvement is validated through the *Weighted Loss* model's *Confusion Matrix* in Figure 2.

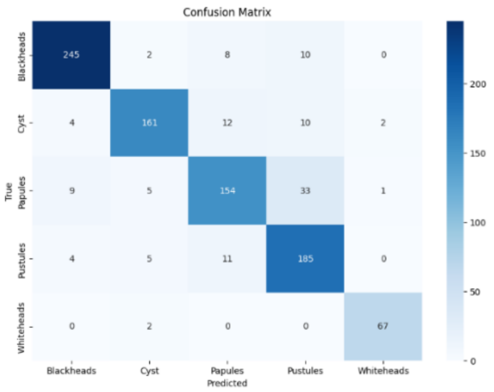

Fig. 5: Confusion Matrix of the Weighted Loss Model

Based on Figure 2, the classification error rate between similar classes was successfully reduced. The error in predicting *papules* as *pustules* decreased from 40 samples (at baseline) to 32 samples. In addition, the classification error *of blackheads* as *pustules* also decreased dramatically from 30 samples to 17 samples. This shows that the model has become more discriminative and "fair" in mapping acne lesion features.

3.4. Comparison and Discussion

A comparative analysis between the two experimental scenarios reveals the fundamental effectiveness of the applied *Cost-Sensitive Learning* strategy. The performance improvement recorded in Table 3 is not merely a statistical fluctuation but reflects a correction to the inductive bias inherent in the *Baseline* model.

Table 3: Comparison of F1-Score Improvement per Class

Class	Baseline (Scenario A)	Weighted (Scenario B)	Improvement
Blackheads	0.86	0.91	+0.05
Cyst	0.81	0.87	+0.06
Papules	0.71	0.81	+0.10
Pustules	0.75	0.81	+0.06
Whiteheads	0.79	0.85	+0.06

3.3.1. Mitigation of Majority Gradient Dominance

In Scenario A, the model tends to ignore minority classes because their loss contribution is drowned out by the large number of majority samples. Conversely, in Scenario B, the massive spike in *F1-Score* in the 'Papules' (+0.10) and 'Whiteheads' (+0.06) classes indicates that re-weighting successfully increased the magnitude of the gradient when the model made errors in those classes. This forces the optimizer to perform more aggressive network weight updates to minimize errors in minority classes, so that visual features previously considered "noise" or unimportant by the *Baseline* model are now learned as crucial discriminative features.

3.3.2. Model Fairness Improvement

Another important finding is the achievement of *Pareto Improvement*, where the performance of minority classes improves without degrading the performance of majority classes. In fact, dominant classes such as 'Blackheads' and 'Cyst' also experienced performance increases of +0.05 and +0.06, respectively. This phenomenon refutes the common concern that *weighted loss* will damage accuracy in majority classes. In fact, by "forcing" the model to distinguish between difficult classes (such as Papules vs. Pustules), the model indirectly learns a more precise and *robust decision boundary* for the entire feature space, which ultimately benefits overall accuracy.

3.3.3. Clinical Implications

From a medical diagnosis perspective, the *Weighted Loss* model offers significantly higher reliability. The increase in *Macro Average F1-Score* from 0.78 to 0.85 indicates that the model has balanced sensitivity across all disease categories. The significant reduction in confusion between inflammatory lesions (papules) and purulent lesions (pustules), as seen in *the Confusion Matrix* (Figure 2 vs. Figure 1), is crucial, because the accuracy of lesion type identification will determine the type of dermatological therapy given to patients. Thus, this proposed method has proven to produce a classification system that is not only mathematically accurate but also clinically valid and safe.

4. Conclusion

Based on the experimental results and discussion, it can be concluded that:

1. The application of the EfficientNetB1 architecture with Weighted Loss Function has proven to be an effective strategic solution for addressing the fundamental constraint of class imbalance in acne image datasets. This approach validates that intervention at the algorithm level can improve the model learning process in skewed data distributions, resulting in models that are more responsive to class variations.
2. The Weighted Cross-Entropy method shows significant performance advantages over the standard method without weighting. This is marked by an increase in the overall accuracy of the model from 80% to 86%, as well as a substantial increase in the Macro F1-Score metric from 0.78 to 0.85, indicating that the model has much better and more consistent generalization capabilities.
3. The proposed model successfully improves sensitivity to minority classes and classes with high visual ambiguity. This success is specifically demonstrated by a 0.10 point increase in the F1-Score for the 'Papules' class (from 0.71 to 0.81). This improvement was achieved without causing a degradation in performance for the majority class, resulting in a more fair, balanced, and reliable diagnostic model for clinical implementation.

References

[1] M. Junayed, M. Islam, A. Jeny, A. Sadeghzadeh, T. Biswas, and A. Shah, "Scarnet: development and validation of a novel deep cnn model for acne scar classification with a new dataset," *Ieee Access*, vol. 10, pp. 1245–1258, 2022, doi: 10.1109/access.2021.3138021.

[2] F. Hu *et al.*, "Advances in intelligent recognition and diagnosis of skin scar images: concepts, methods, challenges, and future trends," *Front Med (Lausanne)*, vol. 12, 2025, doi: 10.3389/fmed.2025.1667087.

[3] Y. Zhang *et al.*, "Ensemble network-based distillation for hyperspectral image classification in the presence of label noise," *Remote Sens (Basel)*, vol. 16, no. 22, p. 4247, 2024, doi: 10.3390/rs16224247.

[4] U. Çinar, R. Çetin-Atalay, and Y. Çetin, "Human hepatocellular carcinoma classification from h&e stained histopathology images with 3d convolutional neural networks and focal loss function," *J Imaging*, vol. 9, no. 2, p. 25, 2023, doi: 10.3390/jimaging9020025.

[5] D. Traini, G. Palmisano, C. Guerriero, and K. Peris, "Artificial intelligence in the assessment and grading of acne vulgaris: a systematic review," *J Pers Med*, vol. 15, no. 6, p. 238, 2025, doi: 10.3390/jpm15060238.

[6] J. Mondragón, E. Rendón, R. Alejo, E. Granda-Gutiérrez, and F. López, "Density-based clustering to deal with highly imbalanced data in multi-class problems," *Mathematics*, vol. 11, no. 18, p. 4008, 2023, doi: 10.3390/math11184008.

[7] M. Najeeb and A. Alariyibi, "Imbalanced dataset effect on cnn-based classifier performance for face recognition," *International Journal of Artificial Intelligence & Applications*, vol. 15, no. 1, pp. 25–41, 2024, doi: 10.5121/ijaia.2024.15102.

[8] Z. Song, Z. Shi, X. Yan, B. Zhang, S. Song, and C. Tang, "An improved weighted cross-entropy-based convolutional neural network for auxiliary diagnosis of pneumonia," *Electronics (Basel)*, vol. 13, no. 15, p. 2929, 2024, doi: 10.3390/electronics13152929.

[9] B. Guo, S. Chen, H. Zhaobin, and G. Xu, "Pattern recognition and analysis: neural network using weighted cross entropy," in *Journal of Physics Conference Series*, 2022, p. 012043. doi: 10.1088/1742-6596/2218/1/012043.

[10] H. Lan *et al.*, "Weakly supervised perivascular spaces segmentation with salient guidance of frangi filter," *Magn Reson Med*, vol. 89, no. 6, pp. 2419–2431, 2023, doi: 10.1002/mrm.29593.

[11] X. Chen *et al.*, "Application of efficientnet-b0 and gru-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions," *Cancer Med*, vol. 12, no. 7, pp. 8690–8699, 2023, doi: 10.1002/cam4.5581.

[12] N. Ujilast, N. Firdausita, C. Aditya, and Y. Azhar, "MRI image-based Alzheimer's disease classification using convolutional neural network: EfficientNet architecture," *Jurnal Resti (Engineering Systems and Information Technology)*, vol. 8, no. 1, pp. 18–25, 2024, doi: 10.29207/resti.v8i1.5457.