ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2024.0482

OPTIMIZING SKIN LESION CLASSIFICATION WITH CONFUSION-AWARE LOSS FUNCTIONS

Qichen Su and Haza Nuzly Abdull Hamed
Faculty of Computing, University Technology Malaysia, Malaysia

Abstract

Early diagnosis of skin cancer is critical to treatment and saving patients' lives, many studies have used Convolutional Neural Networks (CNNs) to achieve this goal. Traditional methods using the Cross Entropy (CE) loss function, however, often struggle with classes that are easily confused, such as Nevus and Melanoma, leading to reduced diagnostic accuracy. To address this, we propose the Confusion-aware Cross Entropy (CCE) loss function, which enhances classification performance by focusing on these easily confused classes. Our method computes the mean of the negative class logits to identify these classes, ensuring the loss calculation prioritizes their accurate classification. Experiments conducted on the publicly available HAM10000 dataset using ResNet50, EfficientNet-B4, Inception-V3, and DenseNet121 demonstrate that our approach significantly outperforms the traditional CE loss function, achieving higher Accuracy, Sensitivity, and Precision. These results underscore the potential of the CCE loss function to improve clinical outcomes by providing more reliable skin lesion classifications.

Keywords:

Skin Lesion Classification, Cross Entropy, Loss Function, Confusion Aware, CNNs

1. INTRODUCTION

. Skin cancer represents a significant and growing health problem. Millions of people are diagnosed with skin cancer every year worldwide [1]. The most dangerous form, melanoma, is particularly alarming due to its high mortality rate if not detected early. Early diagnosis and treatment are paramount for enhancing patient outcomes, minimizing healthcare expenses, and elevating survival rates [2] [3]. Consequently, the accurate and timely classification of skin lesions is a critical component of dermatological practice.

Convolutional neural networks (CNNs) have advanced the development of image analysis in many fields [4], including medical imaging [5-10]. Their ability to automatically learn features from images and achieve high classification accuracy makes them ideally suited for the diagnosis of skin lesions. Dermoscopy, a non-invasive imaging technique, enhances the visualization of subsurface skin structures, providing more detailed and informative images compared to standard clinical photography [11]. This enhanced detail facilitates more accurate diagnoses and is particularly useful for distinguishing between benign and malignant lesions [12].

Despite the success of CNNs, the traditional Cross Entropy (CE) loss function used in training these networks fail in handling classes that are easily confused. The Fig.1 illustrates the confusion matrix of a ResNet50 model in classifying skin lesions using the HAM10000 dataset [13], with an 80% training and 20% testing split. The matrix reveals that the levels of confusion between different classes vary significantly. For example, there is considerable confusion between Actinic Keratoses (AKIEC) and Basal Cell Carcinoma (BCC), as well as between Nevus (NV) and Melanoma (MEL). Such variations in confusion levels underscore the limitations of using the CE loss function, which treats all misclassifications equally and does not account for these differences.

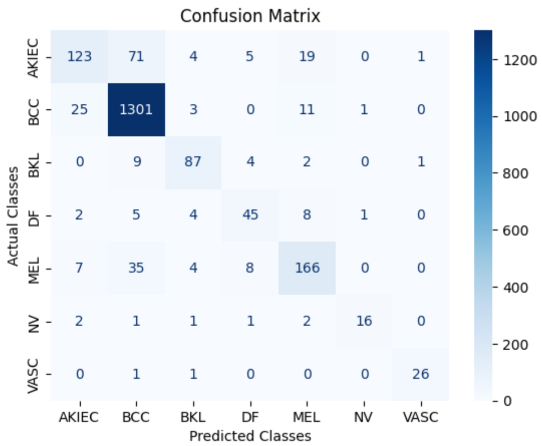

Fig.1. Different confusion level between different type of skin lesions

To address these limitations, we propose the Confusion-aware Cross Entropy (CCE) loss function. The proposed CCE loss enhances the classification performance by giving greater emphasis to easily confused classes. This is achieved by computing the mean of the negative class logits, thereby identifying and prioritizing the accurate classification of these classes. Our approach aims to improve the network's ability to distinguish between similar classes, ultimately leading to more reliable and accurate skin lesion classifications.

In this paper, we present comprehensive experiments conducted on the HAM10000 dataset using various CNN backbones, including ResNet50 [14], EfficientNet-B4 [15], Inception-V3 [16], and DenseNet121 [17]. The results demonstrate that our CCE loss function significantly outperforms the traditional CE loss function, achieving higher Accuracy, Sensitivity, and Precision. These findings highlight the potential of the CCE loss function to enhance clinical decision-making and improve patient outcomes in dermatology.

2. RELATED WORK

The CE loss has been a cornerstone in classification tasks, particularly in the realm of deep learning. It has been extensively used for optimizing models in various domains, including medical image analysis. Despite its effectiveness, several modifications have been proposed to address its limitations and enhance performance.

One notable enhancement is the Focal Loss [18], which applies a modulating factor that emphasizes hard-to-classify examples. This approach has been particularly effective in object detection tasks.

Label smoothing [16], is another significant modification aimed at improving generalization. By softening the targets during training, label smoothing diminishes the model's excessive confidence and contributes to mitigating overfitting.

For medical image classification tasks, the choice of loss function is crucial. A suitable loss function can help the model improve the learning of discriminative features and lead to improved prediction accuracy. For example, [19] introduced a multiscale CNN model for medical image classification that utilized the Mahalanobis distance optimization model as the loss function. The goal of this method is to extract high-quality discriminative features from medical images, which further improves the robustness of the model.

In addition, for the application of machine vision in medical images, the performance of the model can be improved by integrating domain-specific constraints into the loss function. For instance, [20] introduced Feature Centroid Contrast Learning (FCCL) to address domain shift in medical image classification tasks. They employed additional supervision for model training by calculating the contrastive loss between samples and class centers. This approach enhanced classification performance in target domains by mitigating domain shift challenges.

Despite these advancements, the treatment of negative classes in CE loss remains relatively unexplored, particularly for tasks involving fine-grained class distinctions such as skin lesion classification. Our proposed loss function addresses this gap by focusing on the closest negative class, thereby enhancing the model's discriminative capability in scenarios with different confusion levels.

3. METHODOLOGY

The most widely used loss function for CNNs is the CE loss function. Given the logit vector and the ground truth label, CE loss can be written as:

$$CE(\boldsymbol{l}, y) = -\log \frac{\exp(l_y)}{\sum_{c=1}^{C} \exp(l_c)} \tag{1}$$

CE Loss treats all negative classes equally. However, for skin lesion image classification tasks, the confusion level between different classes is not the same, so different negative classes should not be treated equally. Intuitively, negative classes with a high confusion level with the true class should have higher weights in the loss compared to those with a low confusion level. Based on this insight, this study proposes the CCE loss. CCE loss only considers classes that are easily confused with the true class and ignores those that are not, forcing the classification model to learn fine-grained features of the most confusing classes and enhancing its ability to distinguish between easily confused classes.

Specifically, to compute the CCE loss, we first exclude the element corresponding to the true class from the model's output logits, result in

$$\boldsymbol{l}_{\text{neg}} = (l_1, \ldots, l_{y-1}, l_{y+1}, \ldots, l_C) \tag{2}$$

We then calculate the mean of these negative class logits, $\mu_{\text{neg}} = \frac{1}{C-1} \sum_{i \neq y} l_i$. Next, we identify the set of easily confused classes as $\mathrm{H} = \{ j |\ j \neq y \text{ and } l_j > \mu_{\text{neg}} \}$. The CCE loss is then computed by focusing only on these easily confused classes, formulated as

$$\mathrm{CCE}(\boldsymbol{l}, y) = -\log \left(\frac{\exp(l_y)}{\exp(l_y) + \sum_{j \in \mathrm{H}} \exp(l_j)} \right) \tag{3}$$

This approach forces the model to prioritize learning fine-grained distinctions between classes that are similar, thereby enhancing overall classification performance.

4. EXPERIMENTS AND RESULTS

This section presents and analyses the experimental details and results of our proposed CCE loss compared to the CE loss on the ResNet50, EfficientNet-B4, Inception-V3, and DenseNet121 models for skin lesion classification.

4.1 DATASET

For our experiments, we utilized the HAM10000 dataset. It collects dermoscopic images from different populations, which differ in the way they are acquired and stored, thus ensuring the diversity of the dataset. The HAM10000 dataset has broad application prospects in the field of skin diagnosis. It can be used to train machine learning models to achieve automatic diagnosis of pigmented skin lesions. In addition, the dataset can be utilized to evaluate the performance and promote the development of related technologies.

The dataset comprises 10,015 dermoscopic images from different populations acquired and archived over 20 years. Sample images are shown in Fig.2. These images encompass seven distinct diagnostic categories. The sample distribution among various classes in the HAM10000 dataset is imbalanced, with the NV class having the largest number of samples while the MEL class having the smallest, making it a highly challenging dataset. The number of images of each class is shown in Fig.3.

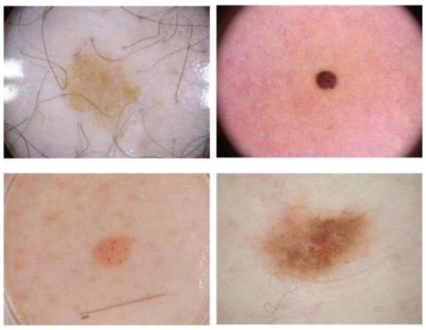

Fig.2. Sample images taken from HAM10000 dataset.

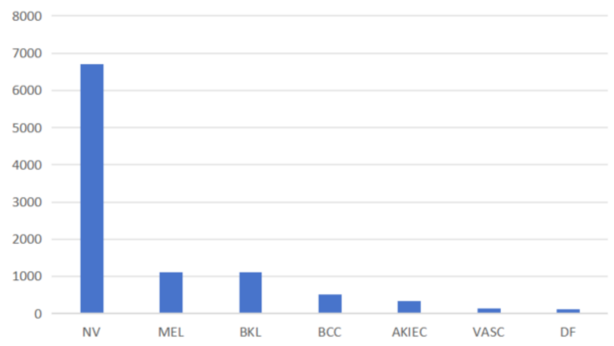

Fig.3. Number of images of each class in HAM10000 dataset.

Each image in the dataset is labelled with the corresponding diagnosis confirmed by histopathology, follow-up examinations, expert consensus, or confirmation by a specialist. This high-quality labelling ensures the reliability of the dataset for training and evaluating machine learning models.

To evaluate the performance of CCE loss function, in this study, we use 80% of the images in the HAM10000 dataset to train the model, and the performance is tested on the remaining 20%. The complexity and diversity of the HAM10000 dataset allows us to rigorously test the ability of our method to accurately classify skin lesions.

4.2 EVALUATION METRICS

To evaluate the performance of our proposed CCE loss function and the baseline models, we employed several standard metrics: Accuracy (ACC), Sensitivity, Precision, and Specificity. These four metrics are each important in the skin disease classification task. Together, they form a comprehensive framework for evaluating model performance, helping researchers understand the performance of the model in different aspects and optimize and improve the model accordingly. These metrics are defined as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}} \tag{4}$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{5}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{6}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{7}$$

By employing these metrics, we can assess the improvements of performance introduced by the proposed CCE loss function compared to the traditional CE loss function. This detailed evaluation is crucial for demonstrating the practical utility of our approach in clinical settings.

Table.1. Performance comparison between different models with CE loss and CCE loss on the HAM10000 dataset.

Method	ACC	Sensitivity	Precision	Specificity
ResNet50 + CE	96.59	77.68	83.63	96.45
ResNet50 + CCE	97.20	80.25	85.69	97.16
EfficientNet-B4 + CE	97.55	84.78	88.09	97.57
EfficientNet-B4 + CCE	97.86	86.65	88.85	97.69
Inception-V3 + CE	90.66	53.36	55.02	93.73
Inception-V3 + CCE	92.41	50.41	63.63	92.96
DenseNet121+ CE	96.84	78.49	85.43	96.76
DenseNet121 + CCE	97.02	82.62	84.70	97.11

4.3 RESULTS

We evaluated the performance of CCE loss function against the traditional CE loss function using four different CNN architectures: ResNet50, EfficientNet-B4, Inception-V3, and DenseNet121. The results are summarized in Table 1.

The ResNet50 model showed an improvement in Accuracy from 96.59% with CE loss to 97.20% with CCE loss. Sensitivity increased from 77.68% to 80.25%, Precision from 83.63% to 85.69%, and Specificity from 96.45% to 97.16%.

EfficientNet-B4 exhibited significant improvements with CCE loss, achieving an Accuracy of 97.86% compared to 97.55% with CE loss. Sensitivity improved from 84.78% to 86.65%, Precision from 88.09% to 88.85%, and Specificity from 97.57% to 97.69%.

The Inception-V3 also benefited from the CCE loss. The Accuracy increased from 90.66% to 92.41%, and Precision showed a notable rise from 55.02% to 63.63%, despite a slight decrease in Sensitivity from 53.36% to 50.41%. Specificity also increased from 93.73% to 92.96%.

The DenseNet121 model showed improvements in Accuracy from 96.84% with CE loss to 97.02% with CCE loss. Sensitivity increased from 78.49% to 82.62%, Precision from 85.43% to 84.70%, and Specificity from 96.76% to 97.11%.

The experimental results conducted on HAM10000 dataset clearly indicate that the proposed CCE loss outperforms the traditional CE loss across all evaluated metrics. By focusing on easily confused classes, the CCE loss helps the model to learn more fine-grained features and make more accurate distinctions between similar classes. The proposed CCE loss is particularly beneficial for skin lesion classification, where the inter-class variation is small.

5. CONCLUSION

This study proposes an enhanced CE loss function, named CCE loss. The rationale behind the CCE loss is rooted in the observation that in skin lesion classification, different classes exhibit varying degrees of confusion. By incorporating the confusion levels directly into the loss computation, the CCE loss encourages the model to focus more on the hard-to-distinguish classes, thereby improving the overall classification accuracy. Future work could explore the application of CCE loss to other medical imaging tasks and further refine the approach to maximize its benefits.

REFERENCES

[1] A. C. Society, "Cancer Facts and Figures 2022", *CA: A Cancer Journal for Clinicians*, Vol. 72, No. 1, pp. 7-33, 2022.

[2] Z. Qin, Z. Liu, P. Zhu and Y. Xue, “A GAN-based Image Synthesis Method for Skin Lesion Classification”, *Computer Methods and Programs in Biomedicine*, Vol. 195, pp. 105568-105576, 2020.

[3] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao and X. Dai, “A GAN-based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification”, *IEEE Access*, Vol. 9, pp. 1-16, 2024.

[4] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects”, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, No. 12, pp. 6999-7019, 2021.

[5] A. Mehmood, Y. Gulzar, Q.M. Ilyas, A. Jabbari, M. Ahmad and S. Iqbal, “SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions”, *Cancers*, Vol. 15, No. 14, pp. 3604-3614, 2023.

[6] S. Maqsood and R. Damasevicius, “Multiclass Skin Lesion Localization and Classification using Deep Learning based Features Fusion and Selection Framework for Smart Healthcare”, *Neural Networks*, Vol. 160, pp. 238-258, 2023.

[7] B. Kalpana, A. Reshmy, S.S. Pandi and S. Dhanasekaran, “OESV-KRF: Optimal Ensemble Support Vector Kernel Random Forest based Early Detection and Classification of Skin Diseases”, *Biomedical Signal Processing and Control*, Vol. 85, pp. 104779-104786, 2023.

[8] E. Goceri, “Classification of Skin Cancer using Adjustable and Fully Convolutional Capsule Layers”, *Biomedical Signal Processing and Control*, Vol. 85, pp. 104949-104958, 2023.

[9] O. Eminaga, “PlexusNet: A Neural Network Architectural Concept for Medical Image Classification”, *Computers in Biology and Medicine*, Vol. 154, pp. 106594-106598, 2023.

[10] V. Anand, S. Gupta, D. Koundal and K. Singh, “Fusion of U-Net and CNN Model for Segmentation and Classification of Skin Lesion from Dermoscopy Images”, *Expert Systems with Applications*, Vol. 213, pp. 119230-119239, 2023.

[11] I.S.A. Abdelhalim, M.F. Mohamed and Y.B. Mahdy, “Data Augmentation for Skin Lesion using Self-Attention based Progressive Generative Adversarial Network”, *Expert Systems with Applications*, Vol. 165, pp. 113922-113934, 2021.

[12] H. Kittler, H. Pehamberger, K. Wolff and M. Binder, “Diagnostic Accuracy of Dermoscopy”, *The Lancet Oncology*, Vol. 3, No. 3, pp. 159-165, 2002.

[13] P.P. Tschandl, C. Rosendahl and H. Kittler, “The HAM10000 Dataset, A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions”, *Scientific Data*, Vol. 5, No. 1, pp. 1-9, 2018.

[14] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition”, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[15] M. Tan and Q. Le, “Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks”, *Proceedings of IEEE International Conference on Machine Learning*, pp. 6105-6114, 2019.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision”, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826, 2016.

[17] G. Huang, Z. Liu, L. Van Der Maaten and K.Q. Weinberger, “Densely Connected Convolutional Networks”, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708, 2017.

[18] T.Y. Lin, P.P. Goyal, R. Girshick, K. He and P.P. Dollar, “Focal Loss for Dense Object Detection”, *Proceedings of IEEE International Conference on Computer Vision*, pp. 2980-2988, 2017.

[19] F. An, X. Li and X. Ma, “Medical Image Classification Algorithm Based on Visual Attention Mechanism-MCNN”, *Oxidative Medicine and Cellular Longevity*, Vol. 2021, No. 1, pp. 1-13, 2021.

[20] W. Zhou, “Contrastive Centroid Supervision Alleviates Domain Shift in Medical Image Classification”, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 816-826, 2022.