International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 13, No. 3, 2026

Research on Dermoscopic Lesion Classification Based on Deformable Transition Layer and Multi-path Hybrid Attention Fusion

Yan Huang [1] and Chaoan Cai [2]

[1] School of Digital Economy and Business, Chongqing College of Mobile Communication, Chongqing 401420, China
[2] School of Information Engineering, Chongqing Electric Power College, Chongqing 400053, China

Abstract: As the largest organ of the human body, the skin undertakes crucial physiological functions such as protecting the organism and regulating metabolism, and its health status is directly related to the quality of human life. Skin lesions are diverse in types and similar in morphology; failure to accurately identify and intervene in a timely manner can easily lead to delayed diagnosis and treatment, posing a serious threat to patients' lives and health. However, dermoscopic images are commonly plagued by irregular lesion morphology, blurred boundaries and low discrimination of fine-grained features, which restrict the classification performance of traditional deep learning models. To address the above pain points, an improved DenseNet121 architecture integrating multi-path hybrid attention and deformable transition layer (HADT-DenseNet) is proposed. Based on DenseNet121, this architecture introduces deformable convolution to replace the traditional convolution in the transition layer to realize adaptive sampling of irregular lesion features. Meanwhile, a multi-path branch structure is designed, with hybrid attention blocks inserted between key dense blocks to enhance attention features in both channel and spatial dimensions. Through a hierarchical feature fusion strategy, the basic dense features are organically combined with attention-enhanced features to improve the multi-level nature of feature expression. Experimental results on the public ISIC 2017 dataset show that HADT-DenseNet achieves a classification accuracy of 92.3%, a precision of 88.2%, an AUC of 96.1% and an F1-score of 91.8%, which is significantly superior to the baseline model DenseNet121 and other improved variants based on common attention mechanisms. This study provides an effective and feasible new approach for the intelligent auxiliary diagnosis of skin lesions.

Keywords: Dermoscopic Image; Lesion Classification; DenseNet121; Deformable Convolution; Hybrid Attention Block.

1. Introduction

As the largest and most intuitive organ of the human body, the skin directly reflects the physiological and pathological conditions of the organism through its health status, and is of great significance for the early warning and clinical diagnosis of diseases. In recent years, the incidence of skin diseases has been rising year by year, becoming a major problem threatening global public health and imposing a heavy burden on the social medical system and patients' quality of life [1]. Among them, skin cancer is a highly prevalent and life-threatening skin disease, and its potential lethality and rising incidence make them an ongoing clinical and preventive challenge worldwide. In the United States alone, there are 3 million new confirmed cases each year [2], and early accurate detection and intervention are the key to reducing the mortality rate of patients with this disease and improving treatment prognosis [3].

At present, dermoscopy is the most commonly used method for the clinical diagnosis of skin lesions. Boasting the advantages of non-invasiveness and convenience, this technology can clearly present the microstructures of the superficial skin and the papillary layer of the dermis, providing an important reference for the differentiation of benign and malignant lesions. However, its practical effect is highly dependent on the clinical professional experience of dermatologists [4]. Nevertheless, lesions in different skin cancer cases are highly similar in morphology, color and texture, yet with subtle distinguishing differences, which makes it easy for inexperienced dermatologists to make misdiagnoses and missed diagnoses, failing to meet the demand for clinically accurate diagnosis [5]. Against this background, the development of efficient and objective intelligent analysis technology for dermoscopic images has become an important breakthrough to address the challenges of clinical diagnosis and improve the efficiency and accuracy of diagnosis.

With the rapid development of deep learning in the field of medical image analysis, convolutional neural network (CNN)-based classification methods for dermoscopic images have achieved remarkable progress. Many scholars have carried out extensive researches on model improvement and obtained good results. For instance, Alhassan and Altmami [6] proposed the DL-PCMNet model, which combines distributed learning, CNN, and Long Short-Term Memory in a parallel structure for skin cancer classification, achieving an accuracy of 97.28% on the ISIC 2019 dataset. Jaehyuk et al. [7] integrated channel-wise and spatial attention mechanisms into four pre-trained CNNs and optimized hyperparameters via Bayesian optimization, where their RegNetX model achieved the highest accuracy of 98.61% for multi-class skin disease classification. Ahmad et al. [8] proposed the SkinDWNet model combined with Gradient Boosting and the SMOTE Tomek method for skin cancer multi-classification, achieving an accuracy of 97.04% on the ISIC 2019 dataset.

Despite the good application potential demonstrated by existing deep learning-based methods, they still face numerous challenges in practical clinical scenarios [9]. On the one hand, lesions in dermoscopic images are often characterized by irregular morphology, blurred boundaries and uneven texture distribution, and the fixed sampling mode of traditional convolutional neural networks is difficult to accurately capture these irregular features [10]. On the other

hand, a single feature extraction path is prone to the loss of key discriminative information, thereby affecting the discriminative ability of classification models and making it difficult to adapt to complex and diverse clinical lesion images.

Among numerous CNN models, DenseNet has been widely applied in the field of dermoscopic image analysis due to its unique dense connection feature, which can effectively alleviate the gradient vanishing problem of deep networks [11] and exhibits excellent stability in medical image classification tasks. However, the ordinary convolution adopted in its transition layer can only achieve fixed grid sampling and channel compression, which cannot adapt to the irregular morphology of lesions and is difficult to fully extract the fine-grained features of lesions. Meanwhile, as an effective means to improve the feature discrimination ability of models, the attention mechanism has been widely used in dermoscopic image classification tasks. Common attention modules such as Squeeze-and-Excitation (SE)[12] and Convolutional Block Attention Module (CBAM)[13] have significantly improved model performance by strengthening key features and suppressing redundant information. Nevertheless, such single-dimensional attention modules are unable to simultaneously capture the global correlation features in both channel and spatial dimensions, resulting in insufficient comprehensiveness of feature extraction.

The hybrid attention block can effectively make up for the deficiencies of single-dimensional attention by fusing channel attention with the Transformer structure and strengthening the weight assignment of key features [14]. However, directly embedding it into the traditional DenseNet architecture is prone to problems such as insufficient feature adaptability and rigid feature fusion, making it difficult to give full play to its advantages. In addition, deformable convolution realizes adaptive sampling of lesion regions by learning offset values, providing an effective means for irregular feature extraction [10]. The idea of combining it with the attention mechanism has become a research direction, but there is still a lack of mature schemes for the efficient fusion of these two mechanisms in the DenseNet architecture while ensuring the accuracy of feature extraction and the stability of the model.

To address the above deficiencies of existing research, and to further improve the accuracy of dermoscopic image classification and adapt to the recognition needs of complex clinical lesions, this paper proposes an improved DenseNet121 architecture integrating multi-path hybrid attention fusion and deformable transition layer, named HADT-DenseNet. The main contributions are as follows: (1). A deformable transition layer is designed, which replaces the ordinary convolution in the transition layer of DenseNet121 with deformable convolution. While maintaining the channel compression function, it realizes adaptive sampling of irregular lesion features and effectively captures the fine-grained morphological features of lesions. (2). A multi-path hybrid attention embedding structure is constructed, with branched hybrid attention blocks set between dense blocks. A hierarchical feature fusion strategy combining basic dense features and attention-enhanced features is adopted to improve the multi-level expression capability of features, and the improvement effect outperforms single attention modules such as SE and CBAM. (3). Systematic experiments are conducted on the ISIC 2017 dataset. Through comparative experiments with common attention-improved models and ablation experiments, the classification performance of HADT-DenseNet is fully verified, providing technical support for the intelligent auxiliary diagnosis of clinical skin lesions.

2. Proposed Method

The proposed HADT-DenseNet architecture takes DenseNet121 as the basic framework, and constructs a complete classification model for dermoscopic lesions through the transformation of deformable transition layer (DT Layer), multi-path hybrid attention block (HAB) embedding and hierarchical feature fusion strategy. The overall process is as follows: the original dermoscopic image is preprocessed and input to the stem layer for initial feature extraction, then processed by the multi-path feature extraction and fusion module, and finally the classification result is output through the convolution layer and linear layer. The overall structure of the architecture is shown in Fig. 1.

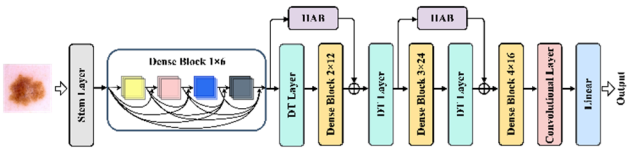


Fig 1. Overall architecture of the HADT-DenseNet model

The core structure of DenseNet121 includes a stem layer, 4 dense blocks and 3 transition layers. The stem layer consists of a 7×7 convolution, a Batch Normalization (BN) layer, a ReLU activation function and a 3×3 max pooling, which converts the input image into an initial feature map. Each dense block contains multiple densely connected convolution layers, realizing dense feature transmission through feature concatenation, and each convolution layer adopts the classic structure of BN-ReLU-Conv. The traditional transition layer is composed of a BN layer, a 1×1 convolution and a 2×2 average pooling, responsible for channel compression and downsampling. HADT-DenseNet retains the core dense connection mechanism of DenseNet121, and focuses on improving the transition layer structure and feature transmission path, introducing deformable convolution and multi-path HABs to enhance the feature extraction and enhancement ability of dermoscopic images.

2.1. Deformable Transition Layer

To address the limitation of fixed sampling in the traditional transition layer, the 1×1 ordinary convolution in the transition layer is replaced with 3×3 deformable convolution to construct a deformable transition layer (DT Layer), which is one of the core improvements of HADT-DenseNet to adapt to the irregular lesion features of dermoscopic images. The layer still maintains the overall structure of BN, convolution layer and average pooling layer, and only optimizes the convolution module with the specific design as follows.

The input of the DT Layer is the output feature map of the previous dense block with the dimension of C×H×W. First, the feature map is normalized through the BN layer to stabilize the training process and alleviate the gradient vanishing problem. Then it is connected to a 3×3 deformable convolution layer, which adds an offset learning branch on the basis of the ordinary 1×1 convolution. The offset branch is composed of a convolution layer with the same kernel size as the current deformable convolution layer, and the output

channel dimension is 2, corresponding to the offset values of a single sampling point in the x and y directions. These offset values are learnable fractional values, updated through backpropagation without additional activation function normalization. As shown in Fig. 2, during the convolution operation, the learned offset values are superimposed on the only central point of the original sampling grid to achieve accurate sampling of irregular regions, thus adapting to the irregular morphological characteristics of dermoscopic lesions.

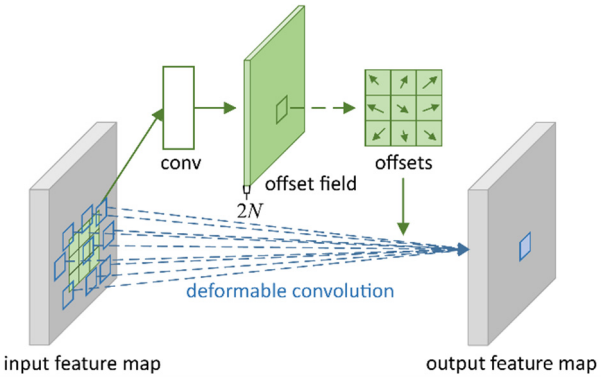


Fig 2. Diagram of deformable convolution [10]

The channel compression rate of the 3×3 deformable convolution maintains the default setting of DenseNet121, that is, the number of output channels is C/2, which ensures the controllability of the number of parameters and computation and avoids model overfitting. Finally, the feature map is downsampled through a 2×2 average pooling layer, reducing the spatial size to 1/2 of the original, with the output dimension of (C/2)×(H/2)×(W/2). The DT Layer not only retains the channel compression and downsampling functions of the traditional transition layer, but also improves the extraction ability of irregular features through adaptive sampling, forming a synergistic effect with the subsequent HAB module.

2.2. Multi-path HAB Embedding and Hierarchical Feature Fusion

To strengthen the weight assignment of key features and solve the problem of incomplete feature capture of single attention modules such as SE and CBAM, HADT-DenseNet designs a multi-path branch structure, inserting HAB between Dense Block 1 and Dense Block 2, as well as between Dense Block 2 and Dense Block 3, forming a parallel feature transmission mode of basic path and attention branch, which is one of the core advantages of the model over common attention-improved models.

The structure of the HAB is designed based on the hybrid attention mechanism, as shown in Figure 3. The module includes two parallel sub-branches: a Channel Attention-based Convolution Block (CAB) and a (Shifted) Window-based Multi-head Self-Attention ((S)W-MSA), and forms a complete feature enhancement unit through residual connection with the subsequent Layer Normalization (LayerNorm) and Multi-Layer Perceptron (MLP).

Among them, the CAB extracts channel features through global average pooling and global max pooling, and generates channel weights through a shared fully connected layer to realize the enhancement of key channels, drawing on the squeeze-excitation idea of the SE module but optimizing the weight generation method. The (S)W-MSA captures the global correlation information of the feature map through window division and multi-head self-attention mechanism, improving the modeling ability of long-distance feature dependence and making up for the deficiency that the spatial attention of the CBAM module can only capture local correlations. In consecutive HABs, Shifted Window Multi-Head Self-Attention (SW-MSA) is adopted at intervals to establish connections between adjacent windows. LayerNorm is used to stabilize the training process, normalizing features before self-attention and MLP operations. The MLP consists of two fully connected layers and an activation function, used to further extract nonlinear features.

The outputs of the two sub-branches are fused by weighted summation, and then residual connected with the input features to obtain the intermediate features of the HAB. Subsequently, the features are processed by LayerNorm and MLP and residual connected again, and finally a feature map with the same dimension as the input is output, ensuring direct fusion with the features of the basic path.

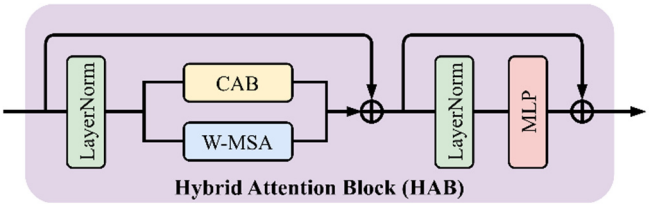


Fig 3. Schematic diagram of the HAB

In HADT-DenseNet, taking the first HAB as an example, feature fusion adopts the element-wise addition method, fusing the output features of the dense block with the output features of the HAB. To realize adaptive feature adjustment, a fusion weight coefficient α is introduced, which is adaptively learned through the training process with a value range of 0 to 1. The fusion formula is as follows:

$$F=\alpha\times F_d+(1-\alpha)\times F_h, \quad (1)$$

where F denotes the final fused feature map generated by the element-wise addition of two feature branches, F_d is the output feature of the dense block, and F_h is the output feature of the HAB. The fused feature map F is input to the subsequent DT Layer for continuous channel compression and adaptive sampling, providing more expressive features for subsequent dense feature extraction.

Combined with the multi-path HAB embedding structure, HADT-DenseNet divides the feature transmission path into three stages and constructs a complete hierarchical feature extraction process, taking into account the integrity of feature extraction and the generalization ability of the model. In the first stage, after the initial feature map is extracted by Dense Block 1, it is divided into two paths: one path undergoes channel compression and downsampling through the DT Layer to retain basic dense features; the other path is input to the HAB for attention enhancement to highlight key lesion features. The output features of the two paths are fused and then input to Dense Block 2, realizing the initial fusion of basic features and attention features. In the second stage, the output features of Dense Block 2 are also divided into two paths, processed by the DT Layer and HAB respectively, then fused and input to Dense Block 3, further enhancing the ability to capture irregular features and key features. In the third stage, the output features of Dense Block 3 are only processed by the DT Layer without setting an HAB branch, avoiding overfitting caused by excessive enhancement of

deep features and reducing the amount of computation at the same time. Finally, deep feature extraction is performed through Dense Block 4, and the classification result is output through the convolution layer and linear layer, completing the entire feature extraction and classification process.

The hierarchical feature transmission path not only retains the advantages of DenseNet in dense feature transmission, but also achieves multi-level feature expression integrating basic features, adaptive sampling features and attention-enhanced features via multi-path fusion. This enhances HADT-DenseNet's ability to classify complex lesions and delivers significantly better performance than improved models employing single attention modules such as SE and CBAM.

2.3. Model Training Parameters

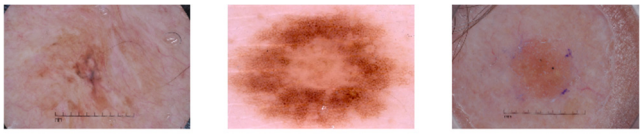
The HADT-DenseNet method proposed in this paper is implemented based on the PyTorch framework, with the input image size adjusted to 224×224. The Adaptive Moment Estimation (Adam) optimization algorithm is adopted as the optimizer, with an initial learning rate set to 0.0001. The learning rate of the deformable convolution offset is set to 0.1 times the basic learning rate to avoid overfitting of the offset. The weight decay coefficient is 0.00001, which suppresses overfitting and improves the generalization ability of the model through L2 regularization. The training batch size is set to 16, and the training epochs are configured to 100. In terms of network structure parameters, HADT-DenseNet contains 4 dense blocks, with the number of layers of each dense block configured as 6, 12, 24 and 16 in sequence, and the feature channel growth rate in the dense blocks is set to 12; the initial value of the feature fusion weight coefficient α is set to 0.5. The cross-entropy loss function is adopted in the experiment to adapt to the classification task of three types of skin lesions.

3. Experiments

3.1. Experimental Setup

3.1.1. Dataset

This paper adopts the public ISIC 2017 dataset, a standard evaluation dataset released by the International Skin Imaging Collaboration (ISIC) in the 2017 Skin Lesion Classification Challenge [15]. The dataset covers three core types of lesions: melanoma (MEL), nevi (NV), and seborrheic keratosis (SK), as shown in Fig. 4. The training set contains 2000 images, including 374 melanoma images, 1372 nevi images and 254 seborrheic keratosis images. The validation set contains 150 images, including 30 melanoma images, 78 nevi images and 42 seborrheic keratosis images. The test set contains 600 images, including 117 melanoma images, 393 nevi images and 90 seborrheic keratosis images. All images are uniformly adjusted to the size of 224×224.



(a) MEL (b) NV (c) SK

Fig 4. Sample images of the ISIC 2017 dataset

3.1.2. Evaluation Metrics

Four types of evaluation metrics are selected to comprehensively evaluate the model performance, covering multiple dimensions such as classification accuracy and generalization ability, with the precision metric added specifically, as follows:

(1). Precision: The proportion of actually positive samples among the samples predicted as positive, used to measure the accuracy of the model's prediction results;

(2). Accuracy: The proportion of all correctly predicted samples to the total samples, used to measure the overall classification effect of the model;

(3). AUC: The area under the Receiver Operating Characteristic curve, with a value range of 0-1. A value closer to 1 indicates a stronger generalization ability and discriminative ability of the model;

(4). F1-score: The harmonic mean of precision and recall, used to comprehensively measure the classification accuracy and recall ability of the model, which effectively alleviates the impact of unbalanced dataset categories.

The four types of evaluation metrics work together to comprehensively reflect the classification effects of HADT-DenseNet and comparison models on different types of lesions, ensuring the reliability and comprehensiveness of the experimental results.

3.2. Experimental Results and Analysis

3.2.1. Comparative Experiment of Different Attention Mechanisms

To explore the influence of different attention mechanisms on the classification performance of DenseNet121, the classification performance of the traditional DenseNet121 and three improved models based on attention mechanisms (DenseNet121+SE, DenseNet121+CBAM, DenseNet121+HAB) are compared, and the experimental results are shown in Table 1. The improvement effects of different attention modules (SE, CBAM, HAB) on the model's precision, accuracy, AUC and F1-score are analyzed emphatically to verify the superiority of the HAB adopted by HADT-DenseNet.

Table 1. Classification performance comparison of different attention-improved models (%)

Model	Precision	Accuracy	AUC	F1-score
DenseNet121(Baseline)	82.5	86.2	91.5	85.7
DenseNet121+SE	84.2	89.3	92.0	88.5
DenseNet121+CBAM	84.7	88.5	92.6	89.0
DenseNet121+HAB	86.6	89.8	93.9	90.3

The experimental results in Table 1 show that after introducing the attention mechanism, all improved models achieve better classification performance than the traditional DenseNet121 baseline model, which fully proves the important role of the attention mechanism in strengthening key features and improving the classification ability of the model, consistent with the conclusions of existing research. Among them, DenseNet121+SE achieves a precision of 84.2%, an accuracy of 89.3%, an AUC of 92.0% and an F1-score of 88.5%, with significant improvements compared with the baseline model, indicating that single channel attention can effectively screen key channel features and

suppress redundant information.

The performance of DenseNet121+CBAM is slightly better than that of DenseNet121+SE, with a precision of 84.7% and an AUC of 92.6%. The reason is that the CBAM module fuses both channel and spatial attention, which can simultaneously strengthen key channel features and key spatial regions, and capture features more comprehensively than the SE module with single channel attention. However, the performance improvement range of both is limited, and the core reason is that SE and CBAM are both simple attention mechanisms with single or dual dimensions, which are difficult to capture the global correlation information of the feature map and have insufficient adaptability to the irregular lesions of dermoscopic images.

DenseNet121+HAB outperforms DenseNet121+SE and DenseNet121+CBAM in all performance indicators, achieving a precision of 86.6%, an accuracy of 89.8%, an AUC of 93.9% and an F1-score of 90.3%. Compared with DenseNet121+CBAM, the precision is increased by 1.9%, the accuracy by 1.3%, the AUC by 1.3%, and the F1-score by 1.3%. The core advantage lies in the HAB adopted by HADT-DenseNet, which fuses channel attention with the Transformer's global correlation capture ability, and can simultaneously strengthen key channel features, key spatial regions and capture long-distance feature dependence. Compared with common attention modules such as SE and CBAM, the feature enhancement effect is more significant, thus achieving the improvement of classification performance, which fully verifies the rationality and superiority of HADT-DenseNet fusing the HAB.

3.2.2. Ablation Experiment Analysis

To verify the effectiveness of each core improved module (HAB, DT Layer) in HADT-DenseNet and clarify the contribution of each module to the model performance, ablation experiments are designed. Four experimental configurations are set in the ablation experiments, namely DenseNet121, DenseNet121+HAB, DenseNet121+DT Layer, and HADT-DenseNet. All experimental configurations adopt the same training parameters and dataset, and the experimental results are shown in Table 2.

Table 2. Ablation experiment results (%)

Model	Precision	Accuracy	AUC	F1-score
DenseNet121 (Baseline)	82.5	86.2	91.5	85.7
DenseNet121+HAB	86.6	89.8	93.9	90.3
DenseNet121+DT Layer	85.7	90.4	94.3	89.6
HADT-DenseNet (Ours)	88.2	92.3	96.1	91.8

The ablation experiment results show that after introducing the multi-path HAB alone, the precision of the model is increased from 82.5% to 86.6%, the accuracy from 86.2% to 89.8%, the AUC to 93.9%, and the F1-score to 90.3%. All performance indicators are significantly improved, and superior to common attention-improved models such as DenseNet121+SE and DenseNet121+CBAM, which further verifies the superiority of the HAB, indicating that the multi-path HAB embedding structure can effectively strengthen key features and improve the feature discriminative ability of the model.

After introducing the DT Layer alone, the precision of the model is increased to 85.7%, the accuracy to 90.4%, and the AUC to 94.3%, which is significantly higher than the results of the baseline. It shows that the adaptive sampling of deformable convolution can effectively capture the irregular lesion features of dermoscopic images, solve the limitation of traditional fixed sampling, and play a significant role in improving the model performance.

The complete HADT-DenseNet model combines the multi-path HAB and the DT Layer, and all performance indicators are further greatly improved, achieving a precision of 88.2%, an accuracy of 92.3%, an AUC of 96.1% and an F1-score of 91.8%. The performance improvement range is more obvious compared with introducing either module alone. This result verifies the synergistic effect of the two core improved modules: the multi-path HAB strengthens the weight assignment of key features and captures global correlation features; the DT Layer realizes adaptive sampling of irregular features and adapts to the morphological characteristics of lesions. The combination of the two realizes multi-level feature expression through the hierarchical fusion strategy, effectively making up for the deficiencies of single module improvement and significantly improving the classification performance of the model, further verifying the rationality and effectiveness of the overall improvement strategy of HADT-DenseNet.

3.2.3. Class-level Classification Results

To further study the performance of the HADT-DenseNet model on different types of skin lesions, experimental analysis is also carried out on the three types of skin lesions in the ISIC 2017 dataset. Fig. 5 shows the classification performance of each of the three types of skin lesions, still measured by four performance indicators: precision, accuracy, AUC and F1-score.

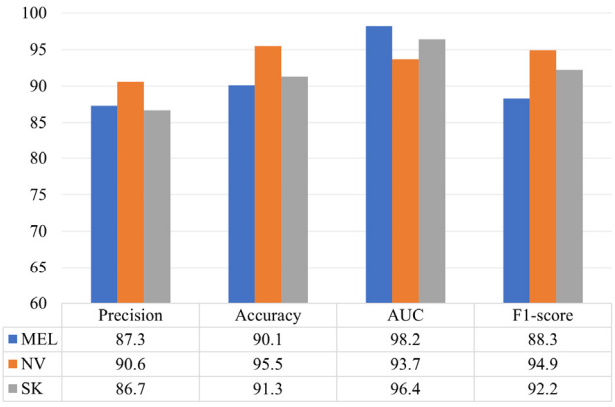


Fig 5. Class-level classification performance comparison (%)

The class-level performance analysis results in Fig. 5 show that HADT-DenseNet achieves a classification precision of 87.3%, an accuracy of 90.1%, an AUC of 98.2% and an F1-score of 88.3% for melanoma, indicating that HADT-DenseNet can effectively capture the irregular features and fine-grained textures of melanoma and improve the discriminative ability of malignant lesions, which is of great clinical significance for dermoscopic auxiliary diagnosis.

Meanwhile, the model achieves a classification precision of 90.6%, an accuracy of 95.5%, an AUC of 93.7% and an F1-score of 94.9% for nevi, and a classification precision of 86.7%, an accuracy of 91.3%, an AUC of 96.4% and an F1-score of 92.2% for seborrheic keratosis, which can effectively avoid the misdiagnosis of benign lesions, further verifying the reliability of HADT-DenseNet.

4. Discussion

The HADT-DenseNet architecture proposed in this paper effectively solves the problems of fixed sampling limitation and incomplete feature capture of single attention in traditional DenseNet121 for dermoscopic lesion classification through the improvement of DT Layer and multi-path HAB fusion. This paper focuses on exploring the influence of different attention mechanisms on model performance and verifies the superiority of the HAB compared with common attention modules such as SE and CBAM. Experimental results show that HADT-DenseNet achieves a classification precision of 88.2%, an accuracy of 92.3%, an AUC of 96.1% and an F1-score of 91.8% on the ISIC 2017 dataset, which is significantly superior to the traditional DenseNet121 and other improved variants based on common attention mechanisms. At the same time, it exhibits excellent generalization and can meet the needs of clinical auxiliary diagnosis.

The core advantage of HADT-DenseNet lies in the realization of dual feature enhancement of adaptive sampling and hybrid attention enhancement, and the full play of the synergistic effect of the two through the multi-path hierarchical fusion strategy, which is also the key to its superiority over common attention-improved models. After replacing the traditional 1×1 convolution with the DT Layer, while maintaining the channel compression function, it can adaptively capture the irregular lesion features and solve the problem that the fixed sampling mode is difficult to adapt to the characteristics of dermoscopic images. The multi-path HAB embedding structure realizes the hierarchical fusion of basic features and attention-enhanced features through parallel branches. The HAB fuses channel attention with the Transformer's global correlation capture ability, and can strengthen key features more comprehensively than single attention modules such as SE and CBAM, improving the multi-level nature of feature expression.

Compared with existing related research, the innovations of this paper are mainly reflected in two aspects. First, deformable convolution is customarily embedded into the transition layer of DenseNet to design a DT Layer, realizing the organic combination of adaptive sampling and channel compression, which is different from the improvement method of only replacing the convolution of the feature extraction layer in existing research. Second, a multi-path HAB embedding and hierarchical fusion structure is constructed, using the hybrid attention mechanism to replace common single attention modules such as SE and CBAM, realizing the multi-level fusion of basic features, adaptive sampling features and attention-enhanced features, and solving the problems of insufficient fusion between attention modules and deformable convolution and incomplete feature capture in existing research.

However, this study still has certain limitations. First, the model is only trained and verified based on the public ISIC 2017 dataset, where the image acquisition equipment and imaging conditions are relatively single. The generalization of the model on clinical images collected by different equipment and under different imaging conditions still needs to be further verified. Second, traditional data augmentation techniques are adopted in the model training process, resulting in insufficient adaptability of the model to low-quality dermoscopic images (such as blurred, low-contrast, and uneven illumination), which are common in clinical applications, and the classification performance of the model on such images needs to be improved.

In view of the above limitations, the future research directions are mainly divided into three aspects. First, expand the dataset scale, include clinical dermoscopic images collected by multi-center and multi-equipment, enrich the diversity of the dataset, cover lesion samples under different imaging conditions and different disease courses, and adopt transfer learning and other technologies to improve the cross-equipment and cross-scenario generalization ability of the model. Second, explore more advanced data augmentation technologies and image restoration technologies to improve the adaptability of the model to low-quality dermoscopic images; optimize the HAB structure at the same time, design lightweight variants, and reduce the amount of computation on the premise of ensuring performance to improve inference efficiency. Third, combine HADT-DenseNet with the clinical diagnosis process to develop an intelligent auxiliary diagnosis system for dermoscopic lesions, realizing functions such as image upload, automatic classification and result visualization, promoting the clinical application of the model, and providing an efficient and accurate auxiliary diagnosis tool for dermatologists.

5. Conclusion

This paper proposes an improved DenseNet121 architecture (HADT-DenseNet) integrating multi-path hybrid attention block and deformable transition layer for the task of dermoscopic lesion classification. It focuses on exploring the influence of different attention mechanisms on model performance and verifies the superiority of the hybrid attention mechanism. Based on DenseNet121, the architecture replaces the ordinary 1×1 convolution in the transition layer with 3×3 deformable convolution to realize adaptive sampling of irregular lesion features; designs a multi-path HAB embedding structure, adopts the hybrid attention mechanism to strengthen key features, which is superior to common attention modules such as SE and CBAM; and adopts a hierarchical feature fusion strategy to organically combine basic features with enhanced features to improve the multi-level nature of feature expression.

Systematic experiments on the public ISIC 2017 dataset show that HADT-DenseNet achieves a classification precision of 88.2%, an accuracy of 92.3%, an AUC of 96.1% and an F1-score of 91.8%, which is significantly superior to the traditional DenseNet121 and other improved variants based on common attention mechanisms. Classification performance comparison experiments clarify the important role of the attention mechanism and prove the superiority of the HAB hybrid attention module over single attention modules such as SE and CBAM; ablation experiments verify the effectiveness of the DT Layer and multi-path HAB, and their synergistic effect can significantly improve the classification performance of the model; class-level performance analysis and generalization verification show that HADT-DenseNet can effectively improve the discriminative ability of malignant lesions with good

generalization and reliability, which can meet the basic needs of clinical auxiliary diagnosis. This study provides a new effective method for the intelligent auxiliary diagnosis of dermoscopic lesions, and has important academic value and practical application prospects. In the future, the performance and clinical applicability of the model can be further improved by expanding the diversity of the dataset, optimizing the model structure and developing an auxiliary diagnosis system, promoting the popularization and application of deep learning technology in the field of dermoscopic diagnosis.

References

[1] Arnold M, Singh D, Laversanne M, et al. Global burden of cutaneous melanoma in 2020 and projections to 2040[J]. JAMA Dermatology. 2022, 158(5): 495-503.

[2] Flohr C, Hay R. Putting the burden of skin diseases on the global map[J]. British Journal of Dermatology. 2021, 184(2): 189-190.

[3] Naqvi S A R, Mobashsher A T, Mohammed B, et al. Benign and malignant skin lesions: Dielectric characterization, modelling and analysis in frequency band 1 to 14 GHz[J]. IEEE Transactions on Biomedical Engineering. 2022, 70(2): 628-639.

[4] Kumari S, Choudhary P K, Shukla R, et al. Recent advances in nanotechnology based combination drug therapy for skin cancer[J]. Journal of Biomaterials Science, Polymer Edition. 2022, 33(11): 1435-1468.

[5] Alshmrani S A, Alotaibi M F, Alfakeeh S A. Fusion of multi CNN features with ANN for early classification of melanoma using dermoscopy images[J]. Discover Sustainability. 2026, 7(1): 211-211.

[6] Alhassan M A, Altmami I N. DL-PCMNet: Distributed Learning Enabled Parallel Convolutional Memory Network for Skin Cancer Classification with Dermatoscopic Images[J]. Diagnostics. 2026, 16(2): 359-359.

[7] Cho J, Shanmugavadivel K, Subramanian M, et al. Attention-aware Deep Learning Models for Dermoscopic Image Classification for Skin Disease Diagnosis[J]. Current Medical Imaging. 2025, 21

[8] Naeem A, Malik H, Din Z M, et al. SkinDWNet: a novel deep learning model for multiclass classification of skin cancers using dermoscopic images[J]. Multimedia Systems. 2025, 31(4): 314-314.

[9] Din Z M, Naeem A, Malik H, et al. Skin cancer classification using a borderline-SMOTE enhanced neural network model on dermoscopic images[J]. Biomedical Signal Processing and Control. 2026, 118: 109691-109691.

[10] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]. Proceedings of the IEEE International Conference on Computer Vision. 2017: 764-773.

[11] Huang G, Liu Z, Laurens V D M, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.

[12] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE. 2018: 7132-7141.

[13] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[C]. Computer Vision-ECCV 2018. Cham: Springer, 2018: 3-19.

[14] Chen X , Wang X , Zhou J , et al. Activating More Pixels in Image Super-Resolution Transformer[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 22367-22377.

[15] Codella N C F, Gutman D, Celebi M E, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)[J]. IEEE Transactions on Medical Imaging, 2018, 37(11): 2642-2659.