

An Attention-Guided Transfer Learning Framework for Automated Basal Cell Carcinoma Diagnosis from Histopathological Images

Alireza Fallahi

Hamedan University of Technology

Soussan Irani

sousanirani@gmail.com

Griffith University

Arash Dehghan

Hamadan University of Medical Sciences

Seied Parsa Saleh

Hamedan University of Technology

Hassan Khotanlou

Bu-Ali Sina University, Hamedan, Iran

Research Article

Keywords: Artificial Intelligence, Basal Cell Carcinoma, Convolutional neural networks, Diagnosis, Histology, Pathology

Posted Date: September 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10703943/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Basal cell carcinoma is the most common form of skin cancer, and accurate histopathological diagnosis remains essential for early treatment and improved clinical outcomes. However, microscopic interpretation is time-consuming and subject to inter-observer variability.

Aim

This study proposes an attention-enhanced transfer learning framework for automated classification of basal cell carcinoma from histopathological images, with emphasis on diagnostic accuracy, computational efficiency, and interpretability.

Methods

A dataset of microscopic color histopathological images containing Basal cell carcinoma and normal skin tissues was utilized. To improve robustness and reduce overfitting under limited data conditions, preprocessing and augmentation techniques including rotation, flipping, cropping, and brightness adjustment were applied. Multiple pre-trained convolutional neural network architectures, including ResNet50, MobileNet, DenseNet121, VGG16, Xception, InceptionV3, and EfficientNet-B2, were fine-tuned using transfer learning. A hybrid channel–spatial attention mechanism was integrated into representative architectures to enhance feature discrimination and provide visual interpretability. The final models were additionally evaluated on an independent external dataset to assess generalizability. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis.

Results

Among the evaluated models, EfficientNet-B2 combined with the attention mechanism, achieved the best overall classification performance, with 98.26% accuracy, 97.35% F1-score and 97.88% AUC, demonstrating a superior balance between accuracy and computational efficiency. Attention-enhanced models consistently outperformed their baseline counterparts, indicating the effectiveness of attention-guided feature learning in histopathological image analysis. EfficientNet-B2 with attention also achieved the highest performance on the primary dataset and maintained 87.54% accuracy, 94.65% F1-score and 86.57% AUC on the independent external dataset. Qualitative attention-map visualization further demonstrated that the proposed framework successfully localized diagnostically relevant tissue regions associated with basal cell carcinoma morphology.

Conclusions

The proposed attention-enhanced transfer learning framework demonstrates strong potential for automated Basal cell carcinoma histopathological classification. The integration of EfficientNet-B2 and hybrid attention mechanisms improves both diagnostic performance and interpretability, supporting the development of clinically applicable and explainable AI-assisted pathology systems.

Background

Basal Cell Carcinoma (BCC) is the most common type of skin cancer, characterized by abnormal and uncontrolled growths within the basal cells of the epidermis. While BCC is highly treatable when detected early, delays in diagnosis can lead to advanced cases that are more challenging to manage [1]. Microscopic examination of tissue biopsies remains the gold standard for diagnosing BCC. This process is labor-intensive, time-consuming, and susceptible to inter-observer variability because it depends on pathologists' subjective interpretations. Moreover, BCC exists in various forms that can resemble other conditions, including molluscum contagiosum, intradermal melanocytic nevus, seborrheic keratosis, trichoepithelioma, melanoma, sebaceous hyperplasia, Bowen's disease, and adnexal carcinoma; hence, it requires distinct differentiation [2].

In recent years, artificial intelligence (AI), particularly deep learning (DL), has shown substantial potential in medical image analysis and digital pathology [3]. Convolutional neural networks (CNNs) have demonstrated strong performance in automating complex image classification and segmentation tasks by learning hierarchical feature representations directly from data [4, 5]. Early studies primarily focused on dermoscopic images, achieving dermatologist-level performance in skin lesion classification [6], demonstrating the feasibility of ResNet-based models for distinguishing benign and malignant cutaneous tumors [7]. These foundational works established deep learning as a powerful tool for dermatological image analysis. Building on these advances, subsequent studies extended DL applications to histopathological and whole-slide images for BCC diagnosis. For instance, Arrastia *et al.* employed a U-Net–based segmentation network on BCC whole-slide images and reported high classification accuracy [8], while Campanella *et al.* introduced a weakly supervised ResNet-based framework achieving near-perfect sensitivity across large pathology datasets [9]. Cao *et al.* further demonstrated the effectiveness of DeepLab V3 with an Xception backbone for BCC analysis [10]. In parallel, classification-focused studies and attention-based models have shown that advanced CNN architectures and attention mechanisms can improve diagnostic accuracy and interpretability [11–13]. Collectively, these works highlight the promise of deep learning for BCC diagnosis while motivating continued investigation into efficient, interpretable, and clinically applicable frameworks.

Despite these advances, limitations remain in the literature. Many approaches focus primarily on segmentation, require extensive whole-slide datasets, or lack interpretability mechanisms tailored to histopathological patterns. Moreover, most classification studies emphasize performance metrics without systematic comparisons between architectures or attention integration. In addition, explicit

evaluation of how attention mechanisms interact with modern, parameter-efficient architectures particularly under limited data conditions typical of clinical repositories is seldom addressed. Also, despite the promising performance reported by previous deep learning studies, many BCC classification systems have been primarily evaluated using internal datasets, making it difficult to determine whether their performance generalizes to images acquired under independent conditions. Recent studies using external cohorts have demonstrated that performance can decline under distribution shifts, highlighting the importance of independent validation for assessing the robustness of computational pathology models [14–17].

In this study, we propose a transfer learning–based framework for the classification of BCC and normal skin tissue using microscopic color images. Multiple state-of-the-art pre-trained CNN architectures, including ResNet50, MobileNet, DenseNet121, VGG16, Xception, InceptionV3, and the parameter-efficient EfficientNet-B2, are systematically evaluated to assess their effectiveness in BCC classification. To further enhance feature discrimination and model interpretability, a hybrid channel–spatial attention mechanism is integrated into selected representative architectures. Comprehensive image preprocessing and data augmentation strategies are employed to improve robustness and generalization under limited data conditions.

Accordingly, the present study not only compares multiple transfer learning architectures and investigates attention-guided feature learning, but also evaluates the trained models on an independent external dataset to assess their generalizability. By combining transfer learning, attention-enhanced feature learning, and parameter-efficient model design, this work aims to provide an accurate, scalable, and clinically applicable solution for automated microscopic BCC diagnosis.

Methods

Dataset

This study utilized a private dataset consisting of 1,150 microscopic color histopathological images from the archive of the pathology department of an educational hospital.

Histologically, BCC appears in different variants, only the most common type, the nodular form, was studied. All images were categorized into two classes: BCC (600 images) and normal skin tissue (550 images). All images were acquired under standard light microscopy with an average resolution of 640 × 480 pixels. Ethical guidelines and data privacy regulations were strictly followed due to the private nature of the dataset.

Images Preprocessing

Prior to model training, all images were resized to match the input size required by each deep learning architecture. To improve generalization and reduce overfitting, extensive data augmentation was applied,

including random rotations, horizontal and vertical flipping, cropping, and brightness adjustments. Through augmentation, the dataset size was expanded to approximately 6,000 images. The augmented dataset was divided into training (80%) and testing (20%) sets. During training, a validation subset was dynamically sampled from the training data to monitor model convergence and prevent overfitting.

Deep Learning Models and Transfer Learning

Multiple convolutional neural network (CNN) architectures pre-trained on the ImageNet dataset were employed using transfer learning, including ResNet50, MobileNet, DenseNet121, VGG16, Xception, InceptionV3, and EfficientNet-B2. These models represent diverse architectural paradigms, residual learning, depthwise separable convolutions, dense connectivity, classical deep CNNs, and parameter-efficient compound scaling, allowing a comprehensive comparison for histopathological image classification. The models were fine-tuned by unfreezing the final fully connected and convolutional layers (18). The evaluation was conducted using two criteria: first, we incrementally unfroze the last n layers (where n could be 15, 20, 25, or 30) to determine the optimal number of layers to unfreeze; second, we fully unfroze all layers and trained the model. It is also important to note that, due to the limited number of layers in MobileNet and VGG16, we did not follow the previous algorithm. Instead, we trained these models by either freezing or fully unfreezing all layers together. This approach allowed us to identify the most effective deep learning model for this study and evaluate the impact of fine-tuning on performance [19]. Models utilized in this study are mentioned below:

ResNet50 leverages residual connections to enable effective training of deep networks and robust feature extraction [20]. MobileNet is a lightweight architecture based on depthwise separable convolutions, offering computational efficiency suitable for resource-constrained settings [21]. DenseNet121 promotes feature reuse through dense connectivity, reducing parameter redundancy while maintaining strong performance on limited datasets [22]. VGG16, a classical CNN with a uniform architecture, provides reliable baseline performance through stacked convolutional layers [23]. Xception enhances efficiency by fully adopting depthwise separable convolutions, enabling improved spatial and channel-wise feature modeling [24]. InceptionV3 captures multiscale features using inception modules, balancing accuracy and computational cost [25]. Finally, EfficientNet-B2 represents a modern, parameter-efficient architecture that employs compound scaling to jointly optimize network depth, width, and input resolution, making it particularly suitable for histopathological image classification tasks with limited data availability.

Attention-Enhanced Network Architecture

To enhance feature discrimination and improve model interpretability, an attention mechanism was integrated into a subset of representative architectures, ResNet50, Xception, and EfficientNet-B2, corresponding to deep residual, depth wise separable, and parameter-efficient CNN designs, respectively. This selective application avoids redundancy while ensuring architectural diversity. The

proposed attention module is a hybrid channel–spatial attention block designed to emphasize diagnostically relevant regions while suppressing background and non-informative tissue structures commonly present in histopathological images [13, 26]. Channel attention models inter-channel dependencies to prioritize informative feature maps, while spatial attention learns a spatial importance map highlighting salient tissue regions [27].

The attention block was inserted after the final convolutional stage of each backbone network, prior to global pooling and classification layers. This placement enables refinement of high-level semantic features while preserving spatial information critical for histopathological interpretation. Figure 1 illustrates the overall attention-enhanced transfer learning pipeline.

Model Fine-Tuning Strategy

Transfer learning was employed by initializing all networks with ImageNet pre-trained weights. To investigate the effect of fine-tuning, multiple training configurations were evaluated:

- (1) all layers frozen,
- (2) partial fine-tuning by unfreezing the last n convolutional layers, and
- (3) full fine-tuning with all trainable layers.

For deeper architectures (ResNet50, DenseNet121, Xception, InceptionV3, and EfficientNet-B2, incremental unfreezing was performed by selectively unfreezing the last 15, 20, 25, and 30 layers. Due to architectural constraints, MobileNet and VGG16 were trained using either fully frozen or fully unfrozen configurations.

Experimental Design and Ablation Study

To evaluate the contribution of the attention mechanism, each selected backbone was trained under two configurations: (i) baseline CNN without attention, and (ii) attention-enhanced CNN. All training settings, data splits, and optimization parameters were kept identical to ensure fair comparison. This design enabled direct assessment of the impact of attention on classification performance and model interpretability.

Experimental Setup and Hyperparameters

To ensure a fair comparison across all architectures and fine-tuning strategies, a standardized training protocol was implemented. All models were developed using the TensorFlow/Keras framework and executed on NVIDIA [RTX 4090/A100] GPUs. We utilized the Adam optimizer due to its adaptive learning rate capabilities, which are particularly effective for transferring learning tasks. For the initial frozen stages, a learning rate of 10^{-3} was applied. During the partial and full fine-tuning phases, the learning

rate was reduced to 10^{-5} to prevent the destruction of pre-trained weights and to ensure stable convergence. Models were trained for a maximum of 50 epochs, incorporating an Early Stopping mechanism. This monitored the validation loss and halted training if no improvement was observed for 10 consecutive epochs (patience = 10), restoring the best-performing weights at the end of the session.

External Dataset and Independent Validation

To assess the generalizability of the developed models beyond the primary dataset, an independent external dataset of histopathological images was used for validation. The external dataset was obtained from the archive of the department of Oral and Maxillofacial Pathology, and comprised 250 images, including 130 BCC images and 120 normal tissue images. None of the images in the external dataset were used for model training, data augmentation, hyperparameter optimization, or model selection. After completion of model development using the primary dataset, the final trained models were directly applied to the external dataset using the same preprocessing pipeline as that used for the primary dataset.

Evaluation Metrics

Model performance was evaluated using accuracy, F1-score, and (Area Under the Curve) AUC. These metrics were selected to provide a comprehensive assessment of classification performance, particularly sensitivity and specificity, which are critical for clinical decision support systems.

Results

Impact of Fine-Tuning Strategy

Figure 2 depicts the accuracy and AUC of various models based on the number of unfrozen layers. The influence of fine-tuning depth was systematically evaluated by progressively unfreezing the final convolutional layers of each network. Across most architectures, partial fine-tuning of the last convolutional blocks resulted in noticeable performance improvements compared to fully frozen models. However, fully unfreezing all layers yielded only marginal gains while significantly increasing computational cost and training time. EfficientNet-B2 and ResNet50 exhibited the highest sensitivity to fine-tuning depth.

Overall Classification Performance

The performance of the evaluated deep learning models was assessed for the binary BCC and normal tissue using accuracy, precision, recall, and F1-score. Experiments were conducted under two configurations: (i) baseline CNN architectures without attention and (ii) attention-enhanced CNNs, following the methodology described in the Method Section. Table 1 summarizes the classification performance of all baseline models under optimal fine-tuning conditions. Among the evaluated architectures, Xception achieved the highest baseline accuracy, followed closely by InceptionV3 and

DenseNet121. EfficientNet-B2 demonstrated competitive performance while maintaining a lower parameter count, highlighting its efficiency for histopathological image classification.

Table 1. Baseline model performance without attention

Model	Accuracy (%)	F1-score (%)	AUC (%)	Optimal model unfrozen layer
ResNet50	96.54	94.32	95.48	30
DenseNet121	97.84	96.36	96.82	25
Xception	94.59	93.59	93.64	25
InceptionV3	96.68	95.75	96.53	30
MobileNet	79.22	78.46	78.22	1
VGG16	87.69	85.64	86.38	0
EfficientNet-B2	98.26	97.35	97.88	15

Effect of Attention Mechanism

To assess the contribution of the attention mechanism, attention-enhanced versions of selected backbone models were evaluated and compared with their baseline counterparts under identical training conditions. Table 2 presents the performance comparison between baseline and attention-enhanced models. The integration of attention consistently improved classification performance across all evaluated architectures, with the most pronounced gains observed in Xception, ResNet50, and EfficientNet-B2.

Table 2. Performance comparison with and without attention mechanism

Model	Accuracy (%)		F1-score (%)		AUC (%)	
	WO-A	W-A	WO-A	W-A	WO-A	W-A
ResNet50	96.54	97.89	94.32	96.28	95.48	96.56
Xception	94.59	96.11	93.59	94.65	93.64	95.36
EfficientNet-B2	98.26	99.49	97.35	99.37	97.88	99.73

* WO-A: Without Attention, W-A: With Attention

The observed improvements suggest that the attention mechanism enables the network to focus on diagnostically relevant tissue regions, thereby enhancing feature representation and classification robustness.

External Dataset Evaluation

Table 3 presents the performance of the evaluated models on the independent external dataset, with and without the attention mechanism. Among the evaluated architectures, EfficientNet-B2 demonstrated the best overall performance. Without the attention mechanism, EfficientNet-B2 achieved an accuracy of 84.94% and an F1-score of 83.58%, with an AUC of 90.82%. Incorporating the attention mechanism improved its performance, increasing the accuracy to 87.54%, the F1-score to 86.57%, and the AUC to 94.65%. These results indicate that the attention-enhanced EfficientNet-B2 model maintained stronger discriminative performance when evaluated on previously unseen data and further support the potential contribution of the attention mechanism to model generalization.

Table 3
Performance of the evaluated models on the external dataset with and without the attention mechanism.

Model	Accuracy (%)		F1-score (%)		AUC (%)	
	WO-A	W-A	WO-A	W-A	WO-A	W-A
ResNet50	80.64	83.36	81.37	83.41	89.59	92.48
Xception	78.47	82.32	79.61	81.34	87.62	90.79
EfficientNet-B2	84.94	87.54	83.58	86.57	90.82	94.65
* WO-A: Without Attention, W-A: With Attention						

Qualitative Attention Analysis

Visual inspection of attention maps revealed that attention-enhanced models consistently highlighted morphologically relevant regions, including tumor nests of BCC (Fig. 3). These findings support the quantitative improvements observed and suggest that the attention mechanism contributes not only to improved accuracy but also to enhanced interpretability of the model’s decision-making process.

Discussion

This study systematically evaluated the performance of multiple pre-trained deep learning architectures to diagnose BCC, and to differentiate between BCC and normal skin tissue using microscopic histopathological images. By combining transfer learning, fine-tuning strategies, and an attention mechanism, the proposed framework achieved high classification accuracy while maintaining interpretability and computational efficiency. Transfer learning was specifically employed to mitigate overfitting associated with limited histopathological datasets. The use of parameter-efficient architectures and augmentation strategies further improved generalization capability under data-constrained conditions.

Among the baseline models, Xception and InceptionV3 demonstrated strong performance, reflecting their ability to capture multiscale and texture-rich features that are characteristic of histopathological images. The superior performance of Xception can be attributed to its use of depthwise separable convolutions, which effectively decouple spatial and channel-wise feature learning, making it particularly

suitable for microscopic tissue analysis. The investigation of fine-tuning depth revealed that partial unfreezing of the final convolutional layers consistently improved model performance across most architectures. In contrast, fully unfreezing all layers resulted in only marginal performance gains while substantially increasing computational cost and training time. This finding highlights the importance of controlled fine-tuning when adapting ImageNet-pretrained models to domain-specific medical imaging tasks. ResNet50 exhibited the highest sensitivity to fine-tuning depth, likely due to its residual connections, which facilitate gradient propagation and enable deeper layers to adapt more effectively to histopathological features. A key contribution of this work is the integration of a hybrid channel–spatial attention mechanism into representative CNN backbones. Attention-enhanced models consistently outperformed their baseline counterparts, with the most pronounced improvements observed for EfficientNet-B2, Xception, and ResNet50. These results suggest that attention mechanisms enable the network to selectively emphasize diagnostically relevant tissue regions, such as tumor nests and peripheral palisading structures, while suppressing irrelevant background information. This selective focus leads to more discriminative feature representations and improved classification robustness.

EfficientNet-B2 achieved the highest overall performance when combined with attention, underscoring the effectiveness of parameter-efficient architectures for histopathological image classification with limited datasets. The compound scaling strategy of EfficientNet, which jointly balances network depth, width, and resolution, allows the model to achieve strong performance without excessive parameter growth. This makes EfficientNet-B2 particularly attractive for clinical and resource-constrained settings. Qualitative analysis of attention maps further supports the quantitative findings by demonstrating that attention-enhanced models produce more localized and interpretable activation patterns. These visual explanations align well with known morphological characteristics of BCC, enhancing trust in the model's decision-making process and supporting its potential clinical utility.

Several previous studies have investigated deep learning–based approaches for skin cancer and BCC classification using either dermoscopic or histopathological images [28–31]. Early studies predominantly relied on conventional CNN architectures such as VGG16 and ResNet, demonstrating promising classification accuracy in skin cancer analysis; however, many of these approaches exhibited limited interpretability, reduced generalizability across datasets, and substantial dependence on large annotated image collections [7, 32–35]. For instance, Esteva *et al.* utilized deep CNNs for skin cancer classification and reported dermatologist-level accuracy; however, the approach relied on large-scale datasets and lacked explicit mechanisms for visual explanation of model decisions (6). Similarly, Han *et al.* applied ResNet-based models for skin lesion classification, focusing primarily on performance metrics without addressing explainability [7]. Similarly, Brinker *et al.* evaluated deep residual CNN architectures for melanoma and non-melanoma skin cancer classification and reported strong diagnostic performance, although model interpretability and cross-dataset validation remained limited [32].

Subsequent studies explored more advanced architectures, including DenseNet and Inception variants, to improve feature representation in histopathological images [36, 37]. Hosny *et al.* demonstrated the

effectiveness of deep CNNs for cancer histopathology analysis but highlighted challenges related to overfitting and limited dataset size [38]. Some works have introduced EfficientNet models due to their parameter efficiency and strong performance. Tan and Le proposed EfficientNet with compound scaling, showing that smaller variants can achieve competitive accuracy with fewer parameters [39]. However, several medical imaging studies have focused on larger EfficientNet variants (e.g., B5 or B7), which increase computational complexity and may not be optimal for limited histopathological datasets [37, 40–42].

Attention mechanisms have also gained traction in medical image analysis to improve both performance and interpretability. Schlemper *et al.* and Oktay *et al.* demonstrated that attention-based models can enhance localization of diagnostically relevant regions in medical images [12, 13]. In skin cancer analysis, attention mechanisms have been shown to improve lesion-focused feature learning [43]; however, a previously published study applied attention to a single backbone architecture without systematic comparison [44]. In contrast to existing literature, the present study integrates a hybrid channel–spatial attention mechanism into multiple representative CNN architectures—ResNet50, Xception, and EfficientNet-B2—and evaluates its impact through a structured ablation framework. The results demonstrate that EfficientNet-B2, combined with attention, achieves superior performance while maintaining a relatively low parameter count, highlighting an advantageous balance between accuracy, efficiency, and interpretability. Additionally, the qualitative attention map analysis provides visual explanations that correspond well with known histopathological characteristics of basal cell carcinoma, thereby enhancing clinical trust and usability.

A key benefit of the proposed attention-enhanced framework, as qualitatively demonstrated in our results, is its ability to generate interpretable visual explanations for its classifications. As illustrated in the accompanying figures, the attention maps derived from our model consistently highlight the diagnostically relevant regions with focusing almost exclusively on the basal cell carcinoma tumors themselves. This focused activation is a significant advancement over standard "black box" deep learning models, as it provides a tangible visual confirmation that the model's decisions are based on the correct pathological features rather than spurious correlations or background artifacts. By directing its focus to the precise areas of interest that a pathologist would examine, the model not only enhances its reliability but also builds a critical bridge of trust with the clinical end-user. This transparency is essential for the clinical adoption of automated diagnostic tools, as it allows practitioners to verify the model's reasoning, thereby increasing confidence in its predictions and facilitating its integration into routine diagnostic workflows for basal cell carcinoma.

The external validation analysis provided an additional assessment of model generalizability beyond the original dataset. Although a reduction in performance compared with the internal evaluation was observed, such a decrease is expected when models are evaluated on data acquired independently of the training distribution. Importantly, the attention-enhanced models maintained comparable performance on the external dataset, suggesting that the attention mechanism may improve the robustness of learned representations rather than merely optimizing performance on the development

dataset. EfficientNet-B2 with attention demonstrated the most favorable overall performance, further supporting the suitability of parameter-efficient architectures combined with attention-guided feature learning for histopathological BCC classification.

Despite the promising results, this study has limitations. The dataset was derived from a single private source, which may limit generalizability. Future work will focus on multi-center validation, extension to multi-class BCC subtypes, and exploration of whole-slide image analysis. Additionally, integrating attention-based explainability with pathologist feedback could further enhance clinical adoption. Overall, compared with prior studies, this work advances the field by jointly addressing performance, computational efficiency, and explainability within a unified framework, making it particularly suitable for real-world histopathological diagnostic applications.

Conclusions

This study presented an attention-enhanced transfer learning framework for the automated classification of basal cell carcinoma (BCC) and normal skin tissue using histopathological images. A comprehensive evaluation of multiple pre-trained CNN architectures demonstrated that EfficientNet-B2 achieved the best overall performance, particularly when combined with the proposed hybrid channel-spatial attention mechanism. The integration of attention improved feature discrimination, enabled localization of diagnostically relevant tissue regions, and enhanced model interpretability. The findings highlight the effectiveness of combining transfer learning, controlled fine-tuning, and attention-guided feature learning for histopathological image analysis under limited data conditions. Furthermore, the attention visualizations provided clinically meaningful explanations that aligned with known morphological characteristics of BCC, supporting greater transparency and trust in AI-assisted diagnosis. Overall, the proposed framework offers a promising, computationally efficient, and interpretable approach for automated BCC detection and may serve as a valuable decision-support tool in digital pathology. Future studies should validate the proposed method using larger multicenter datasets and a broader range of BCC subtypes to further assess its clinical applicability and generalizability.

Declarations

Ethics approval and consent to participate: This study was granted ethical approval by the Ethics Committee of Hamadan University of Medical Sciences (approval No.IR.UMSHA.REC.1404.586).

Consent for publication: Not applicable

Data Availability: Yes, the data that support the findings of this study are available from the corresponding author (S.I.) upon reasonable request.

Competing interests: The authors have no conflicts of interest to declare including financial and non-financial interests.

Funding: This study was financially supported by Hamadan University of Medical Sciences. The grant number was 140407085993.

Authors' contributions:

A. F. collected the data and wrote the manuscript, and constructed the figures and tables.

S. I. Supervisor, contributed to the study design, collected the data and wrote the manuscript.

A. D. & S I. managed the records and archive of images and also collected data.

S P. S. contributed in collecting the data.

H. Kh. a consultant on study design. All authors read and approved the final manuscript.

References

1. Irani S. Pre-cancerous lesions in the oral and maxillofacial region: A literature review with special focus on etiopathogenesis. *Iran J Pathol.* 2016;11(4):303–22. PMID: PMC5563928.
2. Dehghan A, Monsef Esfahani A, Irani S, Jamalpour MR, Sahami A. Basal Cell Carcinoma of the Head and Neck Region: An Analysis of 560 Cases. *Iran J Dermatol.* 2026; In press.
3. Gaffney H, Mirza KM. Pathology in the artificial intelligence era: Guiding innovation and implementation to preserve human insight. *Acad Pathol.* 2025; 28;12(1):100166. 10.1016/j.acpath.2025.100166
4. Dildar M, Akram S, Irfan M, Khan HU, Ramzan M, Mahmood AR, et al. Skin cancer detection: a review using deep learning techniques. *Int J Environ Res Public Health.* 2021;20(10):5479. 10.3390/ijerph18105479.
5. Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S. Skin cancer classification with deep learning: a systematic review. *Front Oncol.* 2022;13:12. 10.3389/fonc.2022.893972.
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017; 2;542(7639):115–118. 10.1038/nature21056
7. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol.* 2018;138(7):1529–38. 10.1016/j.jid.2018.01.028.
8. Le'Clerc Arrastia J, Heilenkötter N, Otero Baguer D, Hauberg-Lotte L, Boskamp T, Hetzer S et al. Deeply supervised UNet for semantic segmentation to assist dermatopathological assessment of basal cell carcinoma. *J Imaging.* 202113;7(4):71. 10.3390/jimaging7040071
9. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019;25(8):1301–9. 10.1038/s41591-019-0508-1.

10. Cao J, Wu J, Zhang JW, Ye JJ, Yu L. Deep learning model trained on mobile phone-acquired frozen section images effectively detects basal cell carcinoma. *arXiv preprint arXiv:201111081*. 2020;18:3023. <https://doi.org/10.48550/arXiv.2011.11081>.
11. Sukanya S, Jerine S. Skin lesion analysis towards melanoma detection using optimized deep learning network. *MTA*. 2023;82(18):27795–817. 10.1007/s11042-023-14454-6.
12. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint*. 2018. <https://doi.org/10.48550/arXiv.1804.03999>. arXiv:180403999.
13. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B et al. Attention gated networks: Learning to leverage salient regions in medical images. *Med Image Anal*. 2019;53:197–207. 10.1016/j.media.2019.01.012
14. Campanella G, Navarrete-Dechent C, Liopyris K, Monnier J, Aleissa S, Minhas B, et al. Deep learning for basal cell carcinoma detection for reflectance confocal microscopy. *J Invest Dermatol*. 2022;142(1):97–103. 10.1016/j.jid.2021.06.015.
15. Geijs DJ, Dooper S, Aswolinskiy W, Hillen LM, Amir AL, Litjens G. Detection and subtyping of basal cell carcinoma in whole-slide histopathology using weakly-supervised learning. *Med Image Anal*. 2024;93:103063. 10.1016/j.media.2023.103063.
16. Björkman J, Lagerroth S, Siarov J, Yacob F, Neittaanmäki N. Enhancing basal cell carcinoma classification in preoperative biopsies via transfer learning with weakly supervised graph transformers. *BMC Med Imaging*. 2025;25(1):166. 10.1186/s12880-025-01710-4.
17. Petzold A, Wessely A, Schliep S, Jiang H, Tran M, Koch EA, et al. Weakly supervised deep learning for cutaneous squamous and basal cell carcinoma in whole-slide histopathology. *J Pathol Clin Res*. 2026;12(2):e70082. 10.1002/2056-4538.70082.
18. Iman M, Arabnia HR, Rasheed K. A review of deep transfer learning and recent advancements. *Technologies*. 2023;11(2):40. <https://doi.org/10.3390/technologies11020040>.
19. Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. *J Big Data*. 2022;9(1):102. 10.1186/s40537-022-00652-w.
20. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778. 10.1109/CVPR.2016.90
21. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*. arXiv:1704.04861. <https://doi.org/10.48550/arXiv.1704.04861>
22. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. arXiv:1608.06993. <https://doi.org/10.48550/arXiv.1608.06993>

23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014. <https://doi.org/10.48550/arXiv.1409.1556>
24. Chollet F, Xception, editors. Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>
25. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. arXiv:1512.00567v3. <https://doi.org/10.48550/arXiv.1512.00567>
26. Woo S, Park J, Lee J-Y, Kweon IS, editors. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV); 2018, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
27. Ilse M, Tomczak J, Welling M, editors. Attention-based deep multiple instance learning. International conference on machine learning; 2018: arXiv:1802.04712v4. <https://doi.org/10.48550/arXiv.1802.04712>
28. Nawaz K, Zanib A, Shabir I, Li J, Wang Y, Mahmood T, et al. Skin cancer detection using dermoscopic images with convolutional neural network. Sci Rep. 2025;15(1):7252. 10.1038/s41598-025-91446-6.
29. Ashfaq N, Suhail Z, Khalid A, Sarwar N, Irshad A, Yaman O, et al. SkinSight: advancing deep learning for skin cancer diagnosis and classification. Discov Comput. 2025;28:63. <https://doi.org/10.1007/s10791-025-09541-1>.
30. Bechelli S, Delhommelle J. Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images. Bioeng (Basel). 2022;27(3):97. 10.3390/bioengineering9030097.
31. Ahmadi Mehr R, Ameri A. Skin Cancer Detection Based on Deep Learning. J Biomed Phys Eng. 2022;12(6):559–68. 10.31661/jbpe.v0i0.2207-1517.
32. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer. 2019;119:11–7. 10.1016/j.ejca.2019.05.023.
33. Magalhaes C, Mendes J, Vardasca R. Systematic review of deep learning techniques in skin cancer detection. BioMedInformatics. 2024;4(4):2251–70. <https://doi.org/10.3390/biomedinformatics4040121>.
34. Furriel BC, Oliveira BD, Prôa R, Paiva JQ, Loureiro RM, Calixto WP, et al. Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review. Front Med (Lausanne). 2024;10:1305954. 10.3389/fmed.2023.1305954.
35. Naqvi M, Gilani SQ, Syed T, Marques O, Kim H-C. Skin cancer detection using deep learning—a review. Diagnostics (Basel). 2023;13(11):1911. 10.3390/diagnostics13111911.
36. Yuan B, Hu Y, Liang Y, Zhu Y, Zhang L, Cai S, et al. Comparative analysis of convolutional neural networks and transformer architectures for breast cancer histopathological image classification. Front Med (Lausanne). 2025;12:1606336. 10.3389/fmed.2025.1606336.

37. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. *Nat Med*. 2020;26(8):1229–34. 10.1038/s41591-020-0942-0.
38. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–10. 10.1038/s41568-018-0016-5.
39. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning; 2019: arXiv:1905.11946v5. <https://doi.org/10.48550/arXiv.1905.11946>
40. Manole I, Butacu A-I, Bejan RN, Tiplica G-S. Enhancing dermatological diagnostics with efficientnet: A deep learning approach. *Bioeng (Basel)*. 2024;11(8):810. 10.3390/bioengineering11080810.
41. Ma X, Shan J, Ning F, Li W, Li H, EFFNet. A skin cancer classification model based on feature fusion and random forests. *PLoS ONE*. 2023;18(10):e0293266. 10.1371/journal.pone.0293266.
42. Rafay A, Hussain W, EfficientSkinDis. An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases. *BSPC*. 2023;85:104869. <https://doi.org/10.1016/j.bspc.2023.104869>.
43. Tai C-eA, Janes E, Czarnecki C, Wong A. Double-condensing attention condenser: Leveraging attention in deep learning to detect skin cancer from skin lesion images. *Sens (Basel)*. 2024;24(22):7231. 10.3390/s24227231.
44. Ding S, Wu Z, Zheng Y, Liu Z, Yang X, Yang X, et al. Deep attention branch networks for skin lesion classification. *Comput Methods Programs Biomed*. 2021;106447. 10.1016/j.cmpb.2021.106447.

Figures

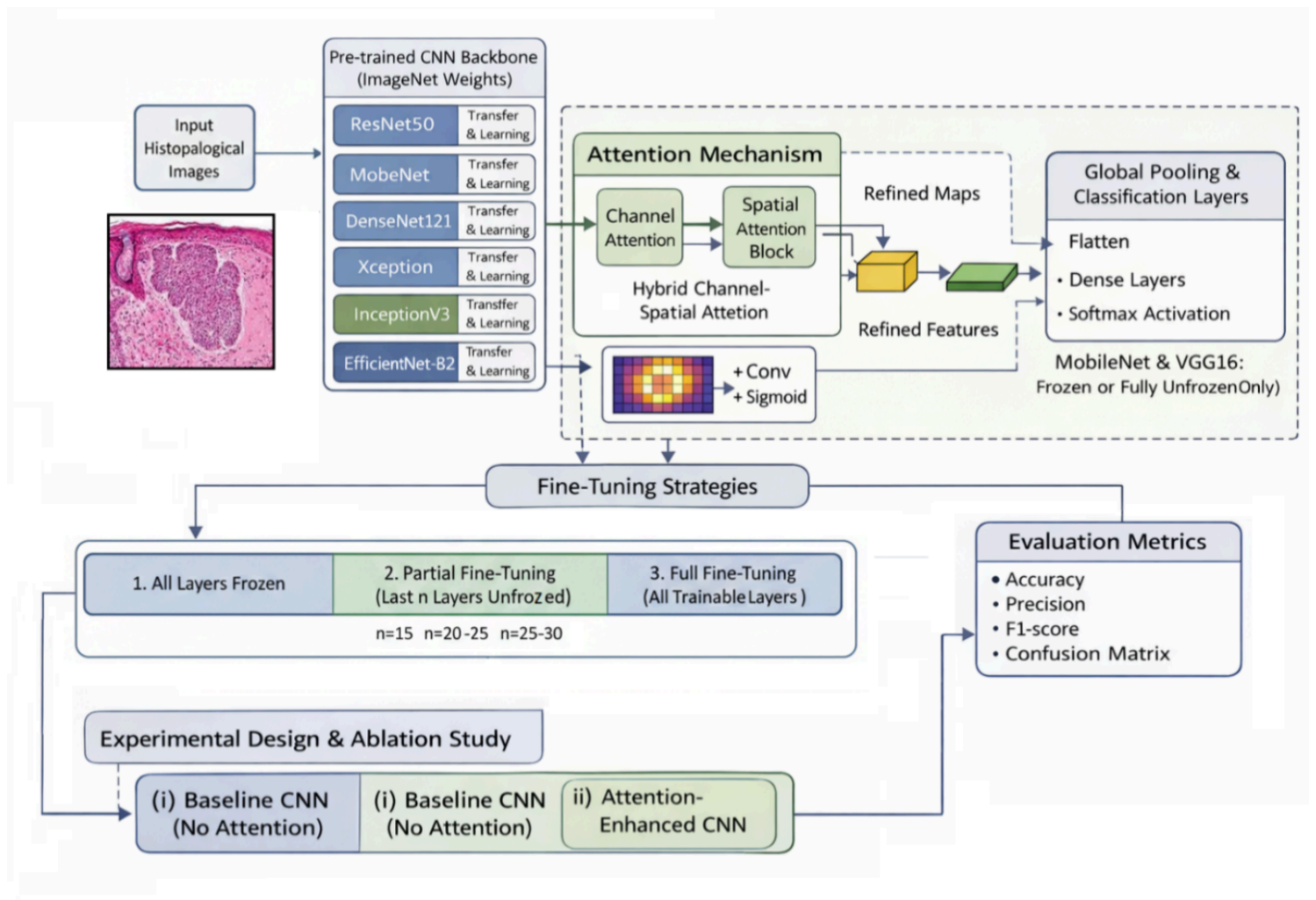


Figure 1

Block Diagram depicting the step-by-step process of the implemented method.

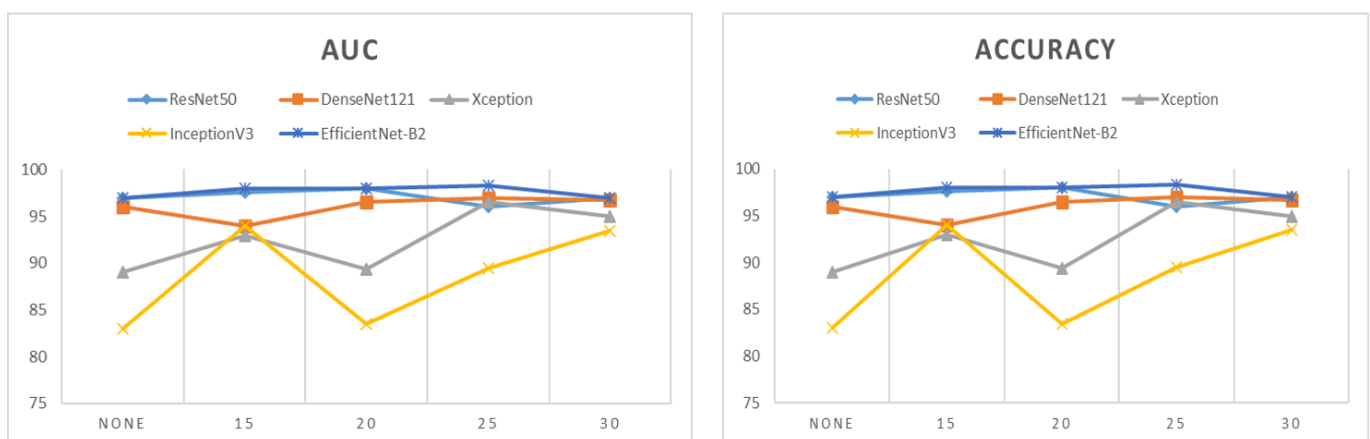


Figure 2

The AUC and accuracy of the models in terms of the number of unfrozen layers.

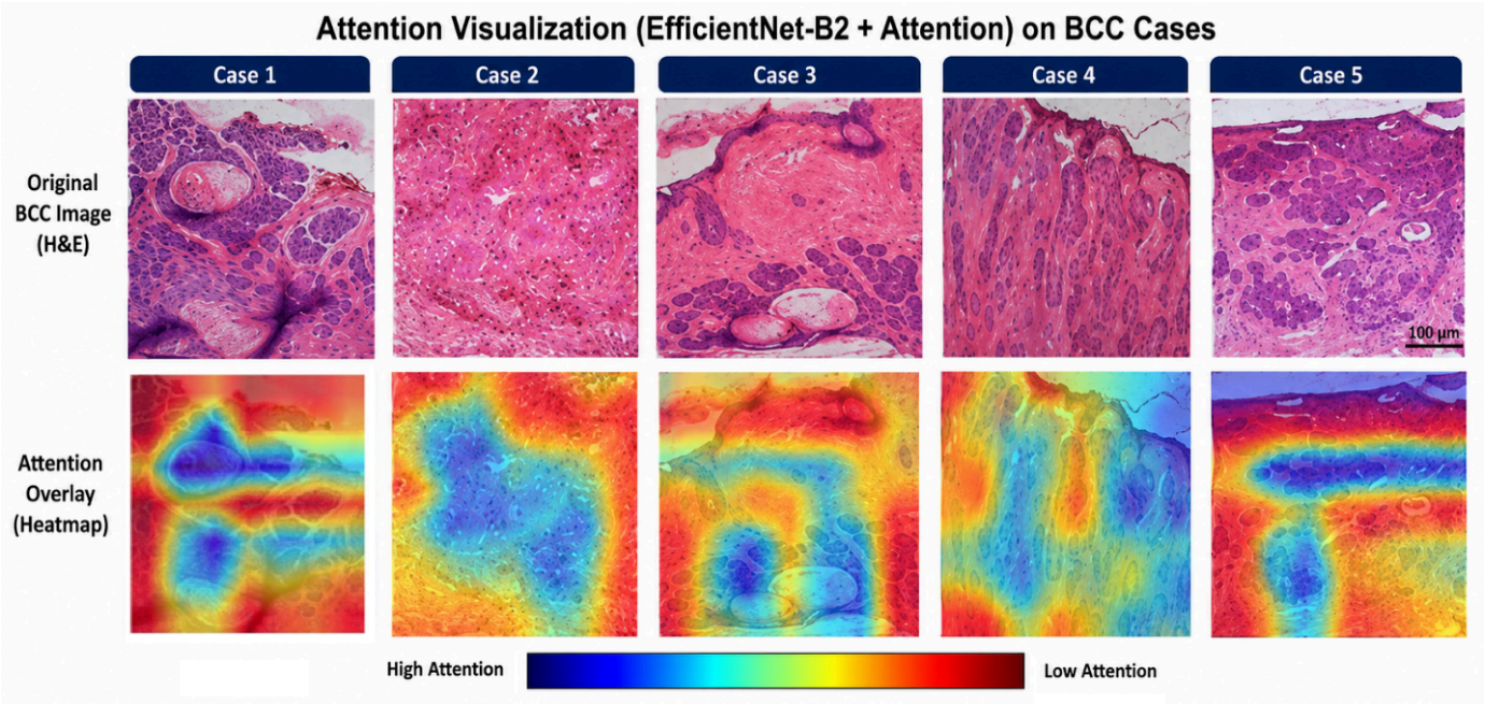


Figure 3

Attention visualization generated by EfficientNet- B2 for five representative Histopathological images of Basal Cell Carcinoma (BCC).