Skin Cancer Lesion Classification Using the HAM10000 Dataset

Linxi Wu
Department of Computer Science
University of North Carolina at Chapel Hill
linxiw@unc.edu

Sabrina Geng
Department of Computer Science
University of North Carolina at Chapel Hill
gkexin@unc.edu

Eddy Liu
Department of Computer Science
University of North Carolina at Chapel Hill
eddyliu5@unc.edu

1 Introduction

Skin cancer poses a global health challenge with increasing incidence rates. Early detection significantly improves survival, but late diagnoses are common, particularly in underserved regions. To address this, AI and machine learning offer promising, cost-effective diagnostic solutions.

This study evaluates two models—Inception-ResNet-v2 with Soft Attention (IRv2-SA) and FixCaps—using the HAM10000 dataset of over 10,000 dermatoscopic images. IRv2-SA employs soft attention for lesion localization, while FixCaps, a capsule network, integrates large-kernel convolutions and attention mechanisms to enhance spatial and feature representation. Data augmentation techniques address class imbalance and variability in lesion appearances.

Results indicate that FixCaps outperforms IRv2-SA in accuracy and generalizability, with fewer parameters and faster training. These findings highlight the potential of advanced AI models for early skin cancer detection, particularly in resource-constrained settings, offering a pathway for more accessible and precise diagnostics.

2 Related Works

Recent studies, such as those by Tyagi et al. (2021) [4], have demonstrated the effectiveness of ViTs in medical image classification tasks like pneumonia detection, where they outperformed traditional CNN-based models. In addition, comparative reviews, such as the one conducted by Mauricio et al. (2023) [5], highlight the growing interest in ViTs for image classification and the advantages they offer in terms of capturing long-range dependencies and context within an image, which could be beneficial for identifying subtle patterns in skin lesions.

Although studies like Tyagi et al. have demonstrated the potential of ViTs in outperforming CNNs in certain tasks like pneumonia detection, their advantages do not necessarily translate to superior performance across all applications. In skin cancer detection, for instance, models incorporating attention mechanisms within CNN architectures, such as FixCaps and Inception-ResNet-V2 with self-attention (IRv2-SA), have shown better sensitivity and accuracy. These hybrid approaches combine the strengths of traditional convolutional layers with attention-based mechanisms, enabling them to outperform standalone ViTs in specific scenarios. While ViTs remain promising, further optimization is required to ensure their performance surpasses or at least matches that of enhanced CNNs in critical applications like skin cancer classification.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

3 Data Preprocessing

The HAM10000 dataset, containing 10,015 dermatoscopic images, was divided into training, validation, and test sets in a 9:1:1 ratio, allocating 8,181 images for training, 1,006 for validation, and 828 for testing. All images were resized to 299×299 pixels to standardize input dimensions, facilitating efficient processing while preserving key features necessary for lesion classification. To tackle the challenge of class imbalance, with malignant lesions like melanoma (MEL) and actinic keratosis (AKIEC) being significantly underrepresented compared to benign lesions such as nevus (NV), data augmentation techniques were employed.

These augmentations included rotations (randomly between -180° and 180°), vertical and horizontal shifts, zooming (in and out), and flipping (horizontal and vertical with a 50 percents probability). This approach simulated real-world image variations, helping the model generalize better by ensuring robustness to orientation, positioning, and scaling changes in lesion images. The augmented dataset increased to 51,629 images, effectively reducing imbalance and improving model diversity. Despite this, accuracy remained stable at 95–96 percents, indicating that the techniques enhanced training data without overfitting, enabling the model to better capture subtle distinctions between malignant and benign lesions. This consistent performance underscores the importance of augmentation in improving data quality and model generalization.

4 Model Structures

4.1 IRv2-SA (Inception-ResNet-v2 with Soft Attention)

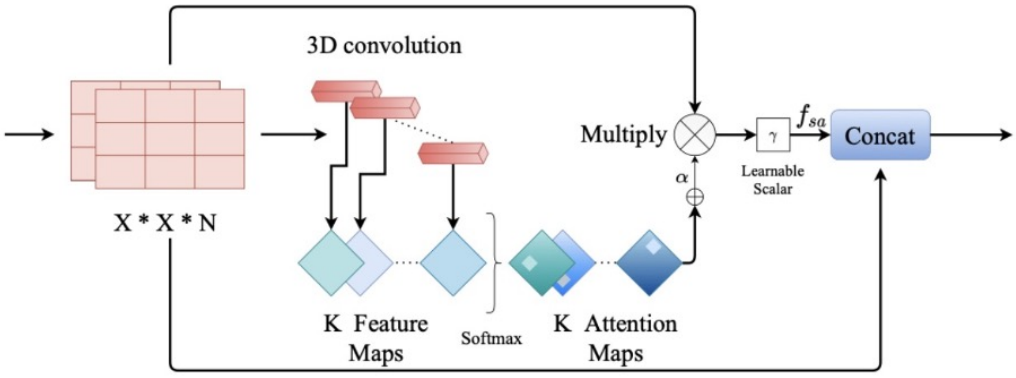

Figure 1: Architecture of IRv2-SA model[6]

IRv2-SA is an enhanced version of the Inception-ResNet-v2 (IRv2) model, combining multi-scale feature extraction and residual connections to handle diverse image characteristics and mitigate the vanishing gradient problem. Its key innovation is a soft attention layer, which improves localization and classification by focusing on the most relevant image regions, such as lesions.

The attention mechanism processes feature maps through 3D convolutions, normalizes them via softmax, and applies weights to highlight significant areas. This enables the model to prioritize lesion-related features while reducing focus on irrelevant background information. In skin cancer detection, this approach enhances the ability to distinguish malignant from benign lesions by identifying subtle morphological differences. By integrating residual connections and soft attention, IRv2-SA achieves robust and efficient performance in complex medical imaging tasks.

4.2 FixCaps (Capsule Network with Improved Features)

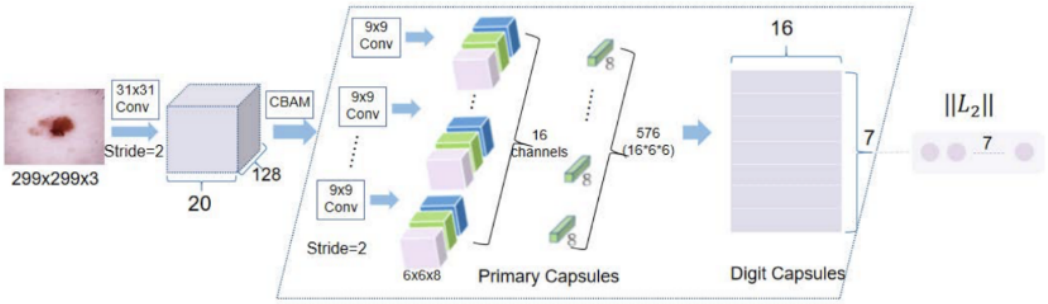

Figure 2: Overview of FixCaps architecture showing the capsule network structure and attention mechanisms [1]

FixCaps is an enhanced capsule network designed to address traditional CNN limitations, such as losing spatial hierarchies during pooling. By grouping neurons into ”capsules” that retain pose, orientation, and spatial relationships, FixCaps preserves detailed spatial information, making it robust to transformations like rotation and scaling.

FixCaps builds on the original idea of capsule networks and introduces several key innovations that enhance its performance for skin cancer detection: 1. Large-Kernel Convolutions (31×31): These significantly expand the network's receptive field, enabling FixCaps to capture broader contextual information and recognize intricate patterns in lesions, which smaller kernels might miss. 2. Convolutional Block Attention Module (CBAM):CBAM enhances focus on relevant features by highlighting key spatial regions and channel dimensions, improving accuracy for lesions with varying shapes, sizes, and locations. 3. Refined Capsule Layers**:** By optimizing the dynamic routing process, FixCaps captures complex relationships between lesion features, improving representation and classification, particularly in distinguishing malignant from benign cases.

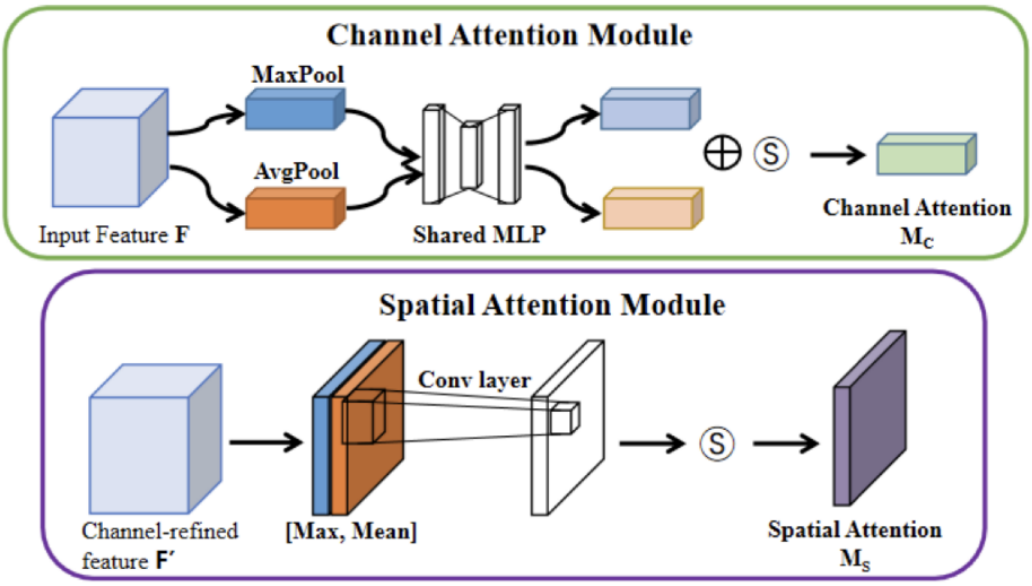

Figure 3: Detailed architecture of FixCaps showing the integration of large-kernel convolutions, CBAM, and refined capsule layers [1]

The structure of three improvements are presented in Figure 3.These advancements make FixCaps highly effective for skin cancer detection, even in challenging scenarios.

5 Results and Discussion

5.1 Evaluation Metrics

Dis.	Recall		F1-score	
	Fix Caps	IRv2-SA	Fix Caps	IRv2-SA
AKIEC	**0.913**	0.520	**0.894**	0.690
BCC	**0.885**	0.880	**0.852**	0.880
BKL	**0.879**	0.830	**0.853**	0.770
DF	**0.667**	0.170	**0.727**	0.290
MEL	0.647	**0.650**	**0.733**	0.660
NV	**0.991**	0.980	**0.990**	0.980
VASC	1.000	**1.000**	0.952	**1.000**

Table 1: Performance comparison across different disease classifications

In this study, metrics like recall and F1-score were used to better assess the models' ability to detect both malignant and benign lesions. FixCaps consistently outperformed IRv2-SA, especially in detecting critical malignant categories like Melanoma (MEL) and Actinic Keratoses (AKIEC), demonstrating high sensitivity essential for early diagnosis. While FixCaps underperformed in the benign Vascular Lesions (VASC) category, this has limited clinical impact, as its strength lies in detecting malignant cases, which are of higher diagnostic priority.

5.2 Impact of Kernel Size

The study highlights the impact of kernel size on FixCaps' performance, with 31×31 convolution kernels significantly improving results. These larger kernels expand the receptive field, allowing the model to capture complex lesion features and global context that smaller kernels (e.g., 3×3) might miss. This enhancement is crucial for detecting lesions with varying shapes and sizes, reinforcing the importance of kernel size in medical image classification tasks where fine details are vital.

6 Conclusion

Skin cancer poses a major public health challenge, emphasizing the need for early and accurate detection. This study evaluated two models, FixCaps and IRv2-SA, for automated diagnosis. FixCaps, with its innovative architecture using large-kernel convolutions and refined capsule layers, outperformed IRv2-SA in accuracy, recall, and F1-scores while maintaining computational efficiency. Its ability to identify subtle lesion differences makes it ideal for resource-constrained clinical settings and mobile applications.

Despite these advances, the HAM10000 dataset's limitations in representing real-world variability highlight the need for diverse datasets and advanced augmentation techniques like GANs. Incorporating explainable AI (XAI) tools can improve transparency and adoption in clinical practice. Future research should explore multimodal data integration and real-time deployment to enhance accessibility and equity in skin cancer diagnostics. These findings mark significant progress toward scalable, AI-driven dermatology solutions, with the potential to ensure early detection is available to all, irrespective of location or resources.

References

[1] Zhangli, L., Songbai, C., Xu, H. & Xinpeng, W. (2022). FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer. https://paperswithcode.com/paper/fixcaps-an-improved-capsules-network-for

[2] Dataverse. (2024). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

[3] Woodman718. (n.d.). FixCaps: LKC architecture image. https://github.com/Woodman718/FixCaps/blob/main/Images/LKC.jpg

[4] Tyagi, K., Pathak, G., Nijhawan, R., & Mittal, A. (2021). Detecting pneumonia using vision transformer and comparing with other techniques. In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 12–16). IEEE. https://doi.org/10.1109/ICECA52323.2021.9676146

[5] Maurício, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences, 13*(9), 5521. https://doi.org/10.3390/app13095521

[6]Datta, S. K., Shaikh, M. A., Srihari, S. N., & Gao, M. (2021). Soft-Attention Improves Skin Cancer Classification Performance. *medRxiv*. https://doi.org/10.1101/2021.05.12.21257114