1 MsBiCNet: Multi-Stage Binary-Cascade CNN for Improving Skin Cancer Detection

Achyut Maheshka, Sanjivanee Jadhav, Pranav Mali, Soham Mhatre, Pramod H. Kachare, Sandeep Sangle
Department of Computer Science and Engineering, RAIT, DY Patil University, Nerul, Navi Mumbai, India.
(email: maheshkaachyut, sanjivaneesj18, malipranav723, soham020804, kachare.pramod1991, sandeepsangle12@gmail.com)

Abstract

Skin cancer's rising incidence highlights the need for accurate detection to ensure effective treatment. Automatic skin cancer detection is challenging due to the visual similarity between benign and malignant lesions, requiring expert evaluation and specialized tools. Limited access to such resources increases the risk of misdiagnosis or delayed treatment. A multi-stage binary-cascade convolution neural network (MsBiCNet) is presented for automatic contactless skin cancer detection using raw images. The proposed MsBiCNet comprises a convolution neural network (CNN) applied sequentially in one-vs-rest stages for multiclass classification. The binary CNN (BiCNet) model has a sequential structure comprising four blocks for feature extraction, three for classification, and a softmax layer for predicting detection probability. Each feature extraction block comprises convolution and max-pooling layers for generating complex feature maps. Each classification block comprises a dense layer followed by a dropout layer to avoid overfitting and improve generalization. The BiCNet (malignant–vs.–benign) and the multiclass MsBiCNet (seven skin lesions) models are trained and evaluated using an open-source human against machine dataset. The results showed accuracies of 87.36% for BiCNet and 80.07% for MsBiCNet, with the MsBiCNet outperforming earlier reported multiclass models. The proposed models enable fast skin lesion diagnosis and may ease healthcare burdens.

Keywords: Convolution neural network, skin cancer detection, machine learning, classification.

Introduction

Skin is the largest organ in our body, consisting of water, proteins, fats, and minerals. It is an essential protective shield against pathogens, physical damage, ultraviolet radiation, and water loss. It regulates body temperature and facilitates excretory functions, vitamin D synthesis, and sensory perception. Structurally, the skin comprises three layers: the epidermis, an outermost layer for protection from environmental damage; the dermis, a middle layer for strength and elasticity; and the hypodermis, an innermost layer for insulation and energy storage. Skin cancer originates in the epidermis layer as the unchecked proliferation of melanocytes that produce the pigment responsible for UV protection. Genetic mutation is one of the dominant sources of this malignancy due to prolonged UV exposure. Fair-skinned individuals who produce lesser melanin are at higher risk than dark-skinned individuals. Skin cancer can be broadly grouped as: basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma (MEL). BCC occurs due to long-term sun exposure and is predominantly observed on surfaces like the face and neck. SCC occurs due to UV exposure and is worsened by exposure to carcinogens or chronic skin diseases. Melanoma (MEL), the deadliest of the three, develops on pre-existing moles, appears as new lesions, and is strongly linked to UV radiation and genetic mutations. Automatic skin cancer detection is challenging due to the visual similarity of lesions, the need for experts, and invasive procedures. Limited resources increase the risk of misdiagnosis or delayed treatment.

Literature review

Several approaches have been reported earlier for automatic skin cancer detection (Naqvi et al., 2023; Bhatt et al., 2023; Dildar et al., 2023; Shah et al., 2023). Some of these approaches using raw images as input are discussed in this section. Thurnhofer-Hemsi and Domínguez (2021)

compared five state-of-the-art convolutional neural networks (CNNs) for skin cancer detection. The comparison using the Human Against Machine (HAM10000) dataset with data augmentation reported the highest accuracy of 89.12%. Ismail et al. (2021) investigated a combination of six pre-trained CNNs for classifying skin lesions, including VGG16, VGG19, ResNet50, Inception V3 DenseNet, and MobileNet. The model recommended combining the three best-pretrained networks to increase classification accuracy. A combination of ResNet50, VGG16, and DenseNet with the insertion of trainable convolution layers reported the highest accuracy of 84.01% on the HAM10000 dataset. Anand et al. (2022) improved VGG16 architecture for early-stage skin cancer diagnosis. The improved model with data augmentation showed an accuracy of 89.09% using a subset of the HAM10000 dataset comprising 5636 augmented images. Ali et al. (2022) investigated variants of EfficientNet for multiclass skin cancer detection. The study identified that intermediate variants (B4 and B5) outperformed deeper variants, with the highest accuracy of 87.91% for EfficientNetB4 using the HAM10000 dataset. Priyadharshini et al. (2023) reported a hybrid extreme learning machine (ML) and teaching-learning-based optimization for melanoma detection. Features were extracted using principal component analysis. The ELM-TLBO reported superior performance with an accuracy of 93.18% using a subset comprising 300 images of melanoma. Mampitiya et al. (2023) compared ResNet50, MobileNet, and six ML models for skin cancer using a subset of the HAM10000 dataset comprising 4694 images. The synthetic minority oversampling technique was used to avoid class imbalance. The comparison showed that SVM outperforms the other algorithms with an accuracy of 99.15% using the PCA dimension reduction. Nugroho et al. (2019) investigated the CNN model with an adaptive moment estimation optimizer for skin cancer using the HAM10000 dataset. The model had two blocks comprising two convolution layers, a max-pooling layer, and two dense layers, with a maximum accuracy of 78%.

The survey of skin cancer detection showed that CNN models are better suited than ML models. The increased complexity of CNN may not increase performance but increases inference time. Most of the earlier studies used a subset of the HAM10000 dataset. Hence, there is a need to develop a lightweight CNN for skin cancer detection in binary and multiclass frameworks.

Proposed skin cancer detection system

The proposed system for skin cancer detection in binary and multiclass framework is shown in Figure 1. The system uses the HAM10000 dataset that is grouped for binary and multiclass detection. It has a preprocessing block for processing images to a fixed size and data augmentation, a model training block for data splitting and binary and multiclass model development using the proposed CNN architecture, and an evaluation block for performance evaluation.

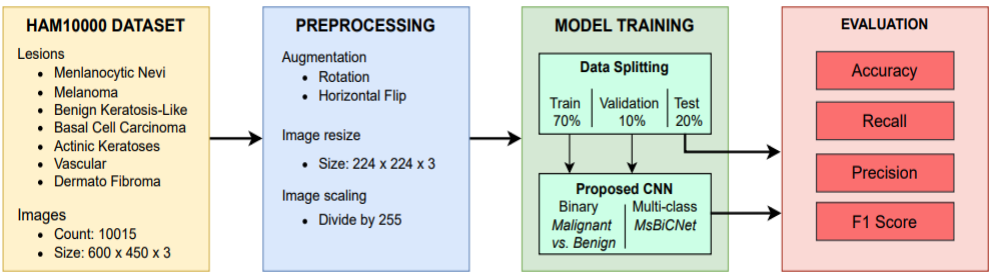


Figure I: Proposed system for skin cancer detection in binary and multiclass framework.

In this work, an open-source HAM10000 dataset is used. It was developed in 2018 to train ML models for classifying skin lesions. The dataset consists of 10015 color images of high-quality dermatoscopic lesions collected from atlases of Dermatology clinical practice. It consists of seven different types of skin lesions. Most of the lesion localizations showed up back in regions of papules or nodules, while a few were seen localized on the lower extremities. Some skin lesions occurred more frequently and were not evenly distributed on anatomical sites. The metadata, such as patient sex, age, and lesion site, is not used in the present work. The skin lesion-wise distribution

of the number of images in the HAM10000 dataset is presented in Table I. This dataset has a class imbalance, variations in image quality, diagnostic methodology variations, and scanner similarity.

Table I: Lesion-wise distribution of the number of images in the HAM10000 dataset.

Skin lesion	Acronym	Training (70%)	Validation (10%)	Testing (20%)	Total
Melanocytic Nevi	NV	4694	670	1341	6705
Melanoma	MEL	779	111	223	1113
Benign Keratosis Like Lesion	BKL	770	109	220	1099
Basal Cell Carcinoma	BCC	360	51	103	514
Actinic Keratosis	AKIEC	230	32	65	327
Vascular Lesion	VASC	100	14	28	142
Dermatofibroma	DF	81	11	23	115
Total	--	7014	998	2003	10015

(Source: Author's compilation)

The ground truth for each image was confirmed by state-of-the-art diagnosis methods like histopathology (53.3%), observations during follow-up (36.9%), dermatologist consensus (9%), and confocal microscopy (0.7%), with histology as the gold standard. For binary detection, seven lesions are grouped as malignant (MEL) and benign (NV, BKL, BCC, AKIEC, VASC, DF). For multiclass detection, seven lesions are grouped into six stages with decreasing images, as shown in Table II. The grouping allows approximate data balance. In the preprocessing block, the raw images are scaled to range 0–1 and resized to a fixed dimension of 224×224×3. Data augmentation comprising horizontal flip and rotation from −10 to 10 degrees is used for generalization during the training. The preprocessed images are split into training (70%) for model building, validation (10%) for generalization, and testing (20%) for evaluation.

Table II: Grouping stages for Multiclass skin cancer detection using HAM10000 dataset.

Stage	Primary lesion	Comparative group
1	NV	MEL, BKL, BCC, AKIEC, VASC, DF
2	MEL	BKL, BCC, AKIEC, VASC, DF
3	BKL	BCC, AKIEC, VASC, DF
4	BCC	AKIEC, VASC, DF
5	AKIEC	VASC, DF
6	VASC	DF

(*Source: Author's compilation*)

Proposed CNN

The proposed skin cancer detection system has binary and multiclass detection models. A CNN architecture is designed for binary classification and is called BiCNet. The BiCNet model has a sequential structure comprising four blocks for feature extraction, three blocks for classification, and a softmax layer for predicting binary detection probability, which is presented in Table III and Figure II. Each feature extraction block comprises convolution and max-pooling layers for generating complex feature maps. The successive convolution layers from input to output have decreasing kernel dimensions and an increasing number of kernels. The first feature extraction block processes the raw input image of (224×224×3) pixels to obtain a feature map of (214×214×32) after the convolution layer and reduces it further to an output feature map of (53×53×32) pixels. The second feature extraction block reduces it further to (23×23×64) pixels.

Similarly, the third and fourth feature extraction blocks generate feature maps of (9×9×128) pixels and (3×3×128) pixels, respectively. The feature extraction part ends with a flatten layer reorganizing the feature maps as an array of 1152 pixels. The classifier part processes the output of the flatten layer to obtain the binary probability for two mutually exhaustive classes. Each classifier block comprises a dense layer followed by a dropout layer to avoid overfitting and improve generalization. The dense layers have a bottleneck structure with two outputs. The dropout layers have decreasing dropout factors of 0.5, 0.3, and 0.2 from input to output. All layers in the BiCNet model use ReLU activation except for the last dense layer with the softmax activation. The model architecture and the number of training parameters are presented in Table III.

Table III: Proposed BiCNet architecture and its training parameters.

Layer	Kernel	Output Shape	Parameters
Input layer	-	(224,224,3)	0
Conv2D_1	(11,11)	(214, 214, 32)	11,648
MaxPool2D_1	(4,4)	(53, 53, 32)	0
Conv2D_2	(7,7)	(47, 47, 64)	100,416
MaxPool2D_2	(2,2)	(23, 23, 64)	0
Conv2D_3	(5,5)	(19, 19, 128)	204,928
MaxPool2D_3	(2,2)	(9, 9, 128)	0
Conv2D_4	(3,3)	(7, 7, 128)	147,584
MaxPool2D_4	(2,2)	(3, 3, 128)	0
Flatten	-	(1152)	0
Dense	-	(128)	147,584
Dropout_1 (0.5)	-	(128)	0
Dense_1	-	(64)	8,256
Dropout_2 (0.3)	-	(64)	0
Dense_2	-	(32)	2,080
Dropout_3 (0.2)	-	(32)	0
Dense_3 (softmax)	-	(2)	66
Total Parameters			**622,562**

(Source: Author's compilation)

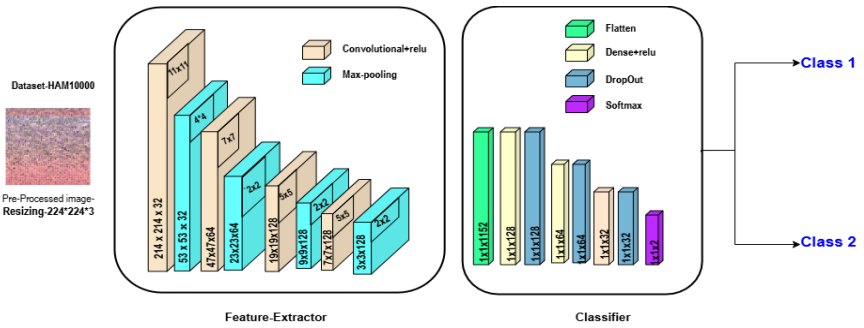


Figure II: Proposed BiCNet model architecture.

For multiclass detection, the BiCNet is trained sequentially in six successive stages, as described in Table II, referred to as MsBiCNet. The training parameters for MSBiCNet are six times those of BiCNet. The models are evaluated, and the results are presented in the next section.

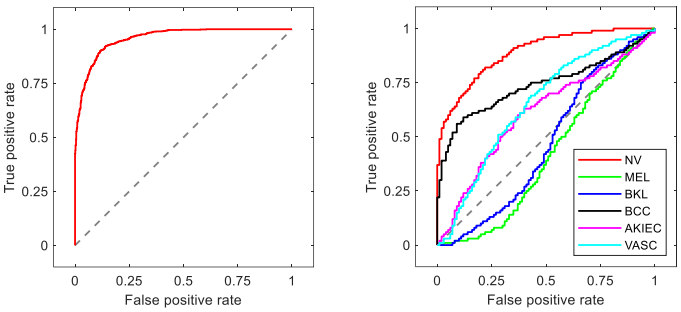
Experimental results

The results for BiCNet and MsBiCNet models are presented in Tables IV–VI and Figure III.

Table IV: Performance evaluation using the BiCNet model for melanoma detection.

Class label	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Malignant (MEL)	88.19	85.35	87.39	86.28
Benign	86.73	88.20	86.40	87.67
Average	87.36	86.78	86.90	86.97

(Source: Author's compilation)



(a) (b)

Figure III: Receiver operating curve for (a) BiCNet and (b) MsBiCNet models.

Table V: Performance evaluation using the MsBiCNet model for multiclass detection.

Stage	Primary lesion	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	NV	82.58	83.53	90.87	87.35
2	MEL	81.12	84.10	86.39	85.23
3	BKL	78.86	76.31	78.91	77.59
4	BCC	66.38	64.31	65.30	70.52
5	AKIEC	77.28	78.57	64.70	70.96
6	VASC	94.20	94.73	95.36	92.67
	Average	80.07	80.26	80.26	80.72

(Source: Author's compilation)

Table VI: Confusion Matrix for the MsBiCNet model using HAM10000 dataset.

		Predicted								
		NV	$\overline{\text{NV}}$		MEL	$\overline{\text{MEL}}$		BKL	$\overline{\text{BKL}}$
Actual	NV	1232	116	MEL	616	732	BKL	1107	439
	$\overline{\text{NV}}$	257	408	$\overline{\text{MEL}}$	428	237	$\overline{\text{BKL}}$	241	226
		BCC	$\overline{\text{BCC}}$		AKIEC	$\overline{\text{AKIEC}}$		VASC	$\overline{\text{VASC}}$
	BCC	335	1013	AKIEC	565	783	VASC	26	26
	$\overline{\text{BCC}}$	272	393	$\overline{\text{AKIEC}}$	155	510	$\overline{\text{VASC}}$	6	59

The results indicate that the BiCNet model has a comparative performance to the stat-of-the-art model, and the MsBiCNet model outperforms state-of-the-art multiclass detection models.

Conclusion

The two frameworks are proposed: BiCNet for binary detection and MsBiCNet for multiclass detection. The BiCNet has a sequence of convolution and max-pooling layers for feature extraction and dense and dropout layers for classification. The BiCNet mode is repeated at six stages to detect each skin lesion separately, referred to as the MsBiCNet model. The BiCNet model showed an accuracy of 87.36% for binary classification (benign vs malignant), and the MsBiCNet model showed an accuracy of 80.07% for multiclass detection (seven skin lesions). The MsBiCNet model outperformed earlier reported multiclass detection models in the literature. In the future, a lightweight architecture should be investigated for skin lesions from different datasets.

References

Ali, K., Shaikh, Z. A., Khan, A. A. and Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets–a first step towards preventing skin cancer. Neuroscience Informatics. 2(4):100034.

Anand, V., Gupta, S., Altameem, A., Nayak, S. R., Poonia, R. C. and Saudagar, A. K. J. (2022). An enhanced transfer learning based classification for diagnosis of skin cancer. Diagnostics. 12(7):1628.

Bhatt, H., Shah, V., Shah, K., Shah, R. and Shah, M. (2023). State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review. Intelligent Medicine. 3(03):180-190.

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O. and Mahnashi, M. H. (2021). Skin cancer detection: a review using deep learning techniques. Int. J. Environmental Res. Public Health. 18(10):5479.

Ghosh, H., Rahat, I. S., Mohanty, S. N., Ravindra, J. V. R. and Sobur, A. (2024). A study on the application of machine learning and deep learning techniques for skin cancer detection. Int. J. Comput. Sys. Engg. 18(1):51-59.

Ismail, M. A., Hameed, N. and Clos, J. (2021). Deep learning-based algorithm for skin cancer classification. In 2020 International Conference on Trends in Computational and Cognitive Engineering (TCCE 2020) (pp. 709-719). Springer Singapore.

Lee, J. R. H., Pavlova, M., Famouri, M. and Wong, A. (2022). Cancer-Net SCa: tailored deep neural network designs for detection of skin cancer from dermoscopy images. BMC Medical Imaging. 22(1):143.

Mampitiya, L. I., Rathnayake, N. and De Silva, S. (2023). Efficient and low-cost skin cancer detection system implementation with a comparative study between traditional and CNN-based models. J. Comput. Cognitive Engg. 2(3):226-235.

Naqvi, M., Gilani, S. Q., Syed, T., Marques, O. and Kim, H. C. (2023). Skin cancer detection using deep learning - a review. Diagnostics. 13(11):1911.

Nugroho, A. A., Slamet, I. and Sugiyanto, S. (2019). Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network. In International Conference on Science and Applied Science (ICSAS) (pp. 020039-1). AIP Publishing.

Priyadharshini, N., Selvanathan, N., Hemalatha, B. and Sureshkumar, C. (2023). A novel hybrid extreme learning machine and teaching–learning-based optimization algorithm for skin cancer detection. Healthcare Analytics. 3:100161.

Shah, A., Shah, M., Pandya, A., Sushra, R., Sushra, R., Mehta, M., Patel, K. and Patel, K. (2023). A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN). Clinical eHealth. 6:76-84.

Thurnhofer-Hemsi, K. and Domínguez, E. (2021). A convolutional neural network framework for accurate skin cancer detection. Neural Process. Lett. 53(5):3073-3093.