IJAR&D
Peer-reviewed | Open Access

International Journal of Academic Research and Development (IJAR&D)
ISSN : 2395-1737 (Print), Volume 12, No. 3, Page No 69-74
DOI : https://doi.org/10.70381/23951737.v12.js3.2026.9
A Quarterly Journal Published by Bharti Publications, New Delhi, India

STAN: A Skin-Tone Adaptive Network for Multi-Branch Dermoscopic Lesion Classification

Pratyay Prasoon
Independent Researchers

Arjun Singh Yadav
Independent Researchers

Abstract

Skin-lesion classification is affected by variation in lesion morphology, image quality, class distribution, and differences in skin appearance. This study presents STAN (Skin-Tone Adaptive Network), a multimodal architecture designed to combine complementary spatial, frequency, and clinical-shape information while conditioning feature fusion on an estimate of skin tone. The model uses a custom multi-scale convolutional network for RGB images, a single-level Haar-wavelet branch for frequency information, and a four-dimensional set of handcrafted lesion descriptors derived from the ABCDE framework. The three representations are projected to a common feature dimension and combined using an Individual Typology Angle (ITA)-conditioned softmax gate. Feature-wise Linear Modulation (FiLM) is then applied to the fused representation. An auxiliary regression head predicts the four handcrafted descriptors during training. The experiment used a five-class subset prepared from the ISIC 2020 image collection: Actinic Keratosis, Basal Cell Carcinoma, Melanoma, Normal Skin, and Squamous Cell Carcinoma. A total of 15,675 images were used. The proposed methodology created a stratified 80:20 split, giving 12,540 training images and 3,135 validation images. The reported validation performance was 90.40% accuracy, 90.32% macro precision, 90.44% macro recall, 90.37% macro F1-score, and 98.81% macro ROC-AUC. The strongest class was Normal Skin, while Actinic Keratosis was the most difficult class. The main errors involved Actinic Keratosis and Basal Cell Carcinoma. These results are encouraging for the proposed multimodal architecture, but they do not establish clinical generalization or the independent contribution of each architectural component. Further work should include an independent test set, repeated runs, ablation experiments, calibration analysis, and subgroup evaluation.

Keywords: *Skin-Lesion Classification, Multimodal Learning, Skin-Tone Adaptation, Individual Typology Angle, FiLM, Haar Wavelet, ABCDE Features, Uncertainty Estimation*

1. Introduction

Skin cancer includes several malignant and precancerous conditions with substantial variation in visual appearance. The appearance of a lesion can change with anatomical location, imaging conditions, lesion morphology, and the patient's skin pigmentation. Deep learning has become a common approach for automated analysis of dermoscopic and clinical images, but high performance on a single dataset does not necessarily translate into reliable performance under changes in acquisition conditions or patient populations.

Most image-classification systems rely primarily on spatial image features. Such representations are useful for capturing color, texture, and lesion morphology, but they do not explicitly separate information according to its frequency characteristics or incorporate clinically motivated shape descriptors. The ABCDE framework—Asymmetry, Border, Color, Diameter, and Evolution—provides a familiar clinical description of pigmented lesions. In the present implementation, four measurable descriptors are used: asymmetry, border irregularity, color variance, and normalized diameter. Evolution is not directly represented because it cannot be obtained from a single static image.

Skin appearance is another source of variation. The present study therefore uses the Individual Typology Angle (ITA) as a quantitative proxy derived from image corner regions assumed to represent surrounding skin. ITA is not treated as a clinical diagnosis or a direct

measurement of Fitzpatrick skin type. Instead, it is used as a conditioning signal that changes the relative contribution of the spatial, frequency, and shape branches.

The main contribution of STAN is the integration of these components into one trainable model. A custom spatial branch extracts multi-scale convolutional features, a Haar-wavelet branch extracts low- and high-frequency information, and an MLP processes the handcrafted descriptors. The three branches are projected to a common 256-dimensional space. An ITA-conditioned gate produces three softmax weights, and the weighted sum is subsequently modulated using FiLM. The model also uses an auxiliary regression objective to retain information about the handcrafted descriptors.

2. Literature Review

Deep learning has produced strong results in automated skin-lesion analysis, with convolutional networks, transfer learning, ensemble methods, and attention-based architectures being widely investigated. Reviews have identified class imbalance, acquisition variability, domain shift, limited diversity in skin pigmentation, and incomplete reproducibility as persistent challenges [1,21,22].

Several studies have investigated CNN-based classification on public skin-lesion datasets. Nawaz et al. used Faster R-CNN and fuzzy k-means clustering for skin-cancer detection, while Khamparia et al. investigated transfer learning in an Internet-of-Health-Things framework [3,4]. Quadir et al. and Garcia examined deep-learning approaches for skin-cancer classification and melanoma detection [5,6]. Other work has investigated optimization, recurrent architectures, transfer learning, and hybrid classical/deep-learning systems [7–15].

Multimodal learning is particularly relevant to STAN because combining complementary information can provide a richer representation than a single image stream. Adebiyi et al. combined lesion images with patient metadata and reported improved performance on HAM10000 and ISIC 2017 [16]. Saeed et al. investigated an adaptive ensemble with metadata on ISIC 2018 and ISIC 2019 and also reported reduced performance during external evaluation [19]. Other studies have examined attention mechanisms and multi-dataset evaluation [18,20]. These results support the need to evaluate architectural ideas under clearly defined data partitions rather than comparing accuracy values across different datasets as though they were directly equivalent.

STAN differs from conventional image-only classifiers by making the fusion mechanism itself dependent on an image-derived skin-tone proxy. The architecture also combines spatial, frequency, and handcrafted clinical-shape information rather than relying on a single pretrained backbone.

3. Materials and Methods

3.1 Dataset Preparation

The study used the ISIC 2020 image collection as the source data. Rather than retaining the complete collection, only the categories required for the experiment were selected and arranged in a folder-per-class structure. The five retained categories were Actinic Keratosis, Basal Cell Carcinoma, Melanoma, Normal Skin, and Squamous Cell Carcinoma. Actinic Keratosis is a precancerous keratinocytic lesion, whereas Basal Cell Carcinoma, Melanoma, and Squamous Cell Carcinoma are malignant skin-lesion categories. Normal Skin was retained as a non-lesion reference class.

The prepared dataset contained 15,675 images. Performed a stratified random split with 20% of the images assigned to validation. The resulting dataset contained 12,540 training images and 3,135 validation images. The validation set was not an independently sourced test set; all reported classification metrics in this study were calculated from this validation partition.

Class	Total	Training	Validation
Actinic Keratosis	3225	2580	645
Basal Cell Carcinoma	3080	2464	616
Melanoma	3160	2528	632
Normal Skin	3315	2652	663
Squamous Cell Carcinoma	2895	2316	579
Total	15675	12540	3135

3.2 Image Processing and Feature Construction

Each image was resized to 260 × 260 pixels. The RGB branch receives the resized image and applies data augmentation during training using horizontal and vertical flips, random rotation, random zoom, random contrast, and random brightness. RGB values are subsequently rescaled by 1/255.

A single-level two-dimensional Haar transform is calculated from a grayscale version of the image. The transform produces four channels: approximation (cA), horizontal detail (cH), vertical detail (cV), and diagonal detail (cD). The resulting representation has spatial dimensions of 130 × 130 with four channels.

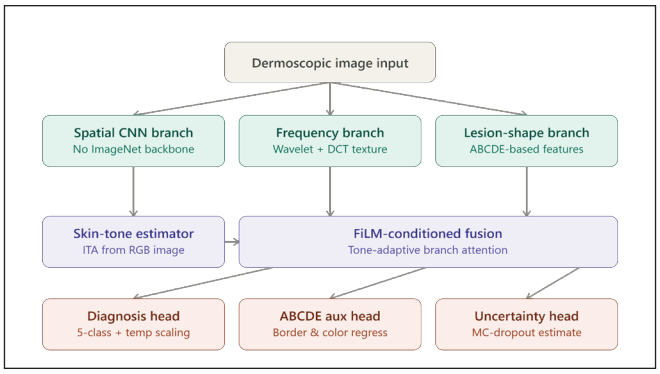
For the skin-tone signal, the image is converted to CIELAB space and four corner patches are sampled. The mean L* and b* values across these patches are used to

calculate ITA as ITA = arctan((L* − 50)/b*) × 180/π. The implementation then divides the resulting angle by 90 and clips the value to approximately [-1, 1]. The four corners are assumed to represent surrounding skin; this assumption can fail when the image framing contains background objects or when the lesion extends into a corner.

The handcrafted branch uses four descriptors. First, asymmetry is estimated from horizontal and vertical reflections of a binary lesion mask. Second, border irregularity is derived from one minus contour circularity. Third, color variance is calculated from the a* and b* components within the segmented lesion. Fourth, normalized diameter is estimated from the minimum enclosing circle relative to the image diagonal. Lesion segmentation uses grayscale conversion, Gaussian smoothing, Otsu thresholding, and selection of the largest external contour. If no contour above the minimum area threshold is found, the four descriptors are set to zero.

3.3 STAN Architecture

STAN contains three parallel feature streams. The spatial stream begins with a 32-channel convolution and four multi-scale blocks. Each block combines standard 3 × 3 convolution, rate-2 dilated 3 × 3 convolution, and 5 × 5 convolution. The concatenated features are batch-normalized, projected using a 1 × 1 convolution, and connected through residual additions where the dimensions match. Global average pooling and a Dense layer produce a 256-dimensional spatial representation.



The frequency stream operates on the four-channel Haar representation. It contains three convolutional stages with batch normalization and max pooling, followed by global average pooling and a Dense layer producing a 96-dimensional representation. The handcrafted branch uses two Dense layers with 32 ReLU units each and produces a 32-dimensional representation.

The three branch outputs are projected to 256 dimensions. The scalar ITA input is passed through a 32-unit hidden layer and a three-unit softmax layer. The three resulting weights correspond to the spatial, frequency, and handcrafted branches. Their weighted sum forms the fused representation.

The fused representation is conditioned using FiLM. ITA is transformed through a 64-unit hidden layer to generate a 256-dimensional scale vector and a 256-dimensional shift vector. The fused feature is multiplied by the scale vector and added to the shift vector. Batch normalization and dropout with a rate of 0.4 follow the FiLM operation.

The diagnosis head contains a Dense(128, ReLU) layer, dropout with rate 0.35, and a five-unit softmax output. The auxiliary head contains Dense(64, ReLU) followed by a four-unit sigmoid output. The total objective is L_total = L_diag + 0.3 L_aux, where L_diag is categorical cross-entropy with label smoothing of 0.1 and L_aux is mean squared error for the four handcrafted targets.

Component	Implementation
Input	260 × 260 × 3 RGB
Spatial branch	Multi-scale CNN, trained from scratch, Dense(256)
Frequency branch	Single-level Haar DWT, 130 × 130 × 4, CNN, Dense(96)
Clinical branch	4 handcrafted descriptors, MLP, Dense(32)
Fusion projection	Each branch projected to 256 dimensions
Adaptive gate	ITA → Dense(32) → Softmax(3)
FiLM	ITA-conditioned γ and β, 256 dimensions
Diagnosis head	Dense(128) → Dropout(0.35) → Softmax(5)
Auxiliary head	Dense(64) → Sigmoid(4)
Total parameters	897,164
Trainable parameters	895,628

3.4 Training Procedure

The network was trained from scratch using Adam. The first training stage used a learning rate of 1×10^{-3} and was configured for up to 25 epochs. A second stage was started with Adam at 1×10^{-4} and was configured for up to 15 epochs. The batch size was 24 and the random seed was set to 42. Class weights were calculated from the training partition and applied to the diagnosis output through per-example sample weights.

Early stopping monitored validation diagnosis accuracy with a patience of six epochs and restored the best weights. The best model was also saved using model checkpointing. ReduceLROnPlateau monitored validation diagnosis loss, reducing the learning rate by a factor of 0.5 after three epochs without improvement, subject to a minimum learning rate of 1×10^{-7}. The first

training stage reached a validation accuracy of 0.80957 at epoch 8 in the recorded run. The second stage later reached a best validation accuracy of approximately 0.904 before early stopping restored the best recorded weights.

3.5 Evaluation

After training, predictions were generated for every batch in the validation dataset. Accuracy, per-class precision, recall, F1-score, macro F1, macro recall, and one-vs-rest macro ROC-AUC were calculated. A 5 × 5 confusion matrix was also generated. Because these metrics were calculated on the same validation partition used for model selection through validation monitoring and checkpointing, they are reported as validation performance rather than independent test performance.

The proposed architecture also contains an optional Monte Carlo dropout function that performs 20 stochastic forward passes and computes mean class probabilities and predictive entropy. This mechanism was implemented as an uncertainty estimate but was not used to change the reported classification metrics.

4. Results

The final recorded validation set contained 3,135 images. STAN achieved 90.40% accuracy, 90.32% macro precision, 90.44% macro recall, 90.37% macro F1-score, and 98.81% macro ROC-AUC.

Class	Precision	Recall	F1-score	Support
Actinic Keratosis	0.824	0.794	0.809	645
Basal Cell Carcinoma	0.852	0.860	0.856	616
Melanoma	0.889	0.896	0.892	632
Normal Skin	0.998	1.000	0.999	663
Squamous Cell Carcinoma	0.953	0.972	0.962	579
Macro average	0.903	0.904	0.904	3135

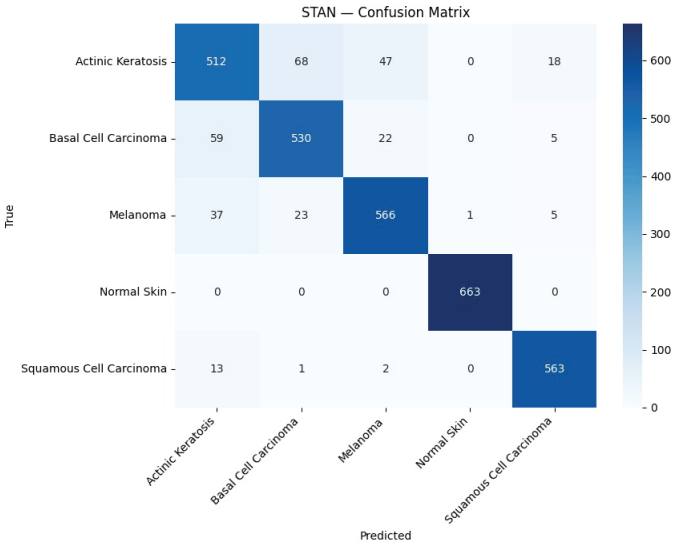
4.1 Confusion Matrix Analysis

The confusion matrix contained 2,834 correct predictions out of 3,135 validation images, corresponding to 90.40% accuracy. Actinic Keratosis had the lowest recall, with 512 of 645 images correctly classified. The largest error flow was from Actinic Keratosis to Basal Cell Carcinoma, with 68 images assigned to BCC. A further 47 Actinic Keratosis images were classified as Melanoma. Basal Cell Carcinoma produced 59 errors toward Actinic Keratosis and 22 toward Melanoma. Melanoma produced 37 errors toward Actinic Keratosis and 23 toward Basal Cell Carcinoma.

Normal Skin had 663 correct predictions out of 663 validation examples. Squamous Cell Carcinoma achieved the highest recall among the malignant classes, with 563 of 579 images correctly classified. The concentration of errors among Actinic Keratosis, Basal Cell Carcinoma, and Melanoma indicates that the remaining errors are not uniformly distributed across classes.

4.2 Interpretation

The validation results indicate that the proposed architecture can learn useful representations from the combination of image, frequency, and handcrafted inputs. The high macro ROC-AUC suggests that the model produced generally useful class probability rankings, while the lower accuracy and macro F1 show that some class boundaries remain difficult. In particular, Actinic Keratosis was substantially harder to classify than Normal Skin or Squamous Cell Carcinoma.



Because no component-level ablation study was performed, the individual contribution of ITA-conditioned gating, FiLM, the wavelet branch, and the auxiliary loss cannot be isolated.

4.3 Comparison with Previous Work

Published skin-lesion results are difficult to compare directly because studies use different datasets, class definitions, preprocessing pipelines, data partitions, and evaluation protocols. For that reason, the following comparisons are contextual rather than claims of superiority.

Study	Dataset/task	Accuracy (%)	F1 / other reported	ROC-AUC
Fraiwan & Faouri [10]	HAM10000, 7 classes	82.90	F1 74.40	NR
Wang et al. [13]	ISIC-2019, binary	90.67	NR	>0.93
Tembhurne et al. [12]	ISIC-derived, binary	93.00	Recall: 99.7% benign / 86% malignant	NR
Adebiyi et al. [16]	HAM10000	94.11	NR	94.26
Saeed et al. [19]	ISIC-2018	93.20	P/R/F1: 93/93/93	97.30
Saeed et al. [19]	ISIC-2019	91.10	P/R/F1: 92/93/92	95.50
STAN (this study)	ISIC 2020, 5 classes; 3,135 validation images	90.40	Macro F1 90.37	98.81

The STAN result is within the range reported in previous skin-lesion studies, but the comparison should not be interpreted as a benchmark ranking. The primary purpose of the experiment was to evaluate the proposed multimodal architecture under the selected five-class setup.

5. Conclusion

This study developed and evaluated STAN, a multimodal skin-lesion classifier that combines a custom spatial CNN, Haar-wavelet features, handcrafted lesion descriptors, ITA-conditioned branch gating, and FiLM conditioning. The implementation also uses an auxiliary regression objective to supervise the handcrafted representation.

Using 15,675 images from a five-class subset of the ISIC 2020 collection, the recorded experiment used a stratified 80:20 split and achieved 90.40% validation accuracy, 90.37% macro F1-score, and 98.81% macro ROC-AUC. The main classification difficulty was concentrated in Actinic Keratosis, particularly its confusion with Basal Cell Carcinoma and Melanoma.

Acknowledgement

The authors acknowledge the organizations responsible for the public skin-lesion data used in this study and the open-source software ecosystem used to implement the model.

Declarations

Competing interests: The authors declare no competing interests.

Funding: No external funding was used for this study.

Ethics: The study used publicly available image data and involved no direct recruitment or intervention. Dataset-specific governance and licensing requirements should be recorded with the final dataset citation.

Data availability: The study used a selected five-class subset prepared from the ISIC 2020 image collection.

References

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. International Journal of Environmental Research and Public Health, 18(10), 5479. https://doi.org/10.3390/ijerph18105479

Rezk, E., Eltorki, M., & El-Dakhakhni, W. (2022). Improving Skin Color Diversity in Cancer Detection: Deep Learning Approach. JMIR Dermatology, 5(3), e39143. https://doi.org/10.2196/39143

Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. Microscopy Research and Technique, 85(1), 339–351. https://doi.org/10.1002/jemt.23908

Khamparia, A., Singh, P. K., Rani, P., Samanta, D., Khanna, A., & Bhushan, B. (2021). An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. Transactions on Emerging Telecommunications Technologies. https://doi.org/10.1002/ett.3963

S., R., J., P., Quadir Md, A., Jackson J, C., Sharma, S., & B., R. (2022). Skin Cancer Detection using Deep Learning. Research Journal of Pharmacy and Technology, 15(10), 4519–4525. https://doi.org/10.52711/0974-360X.2022.00758

Garcia, S. I. (2021). Meta-learning for skin cancer detection using Deep Learning Techniques. arXiv. https://doi.org/10.48550/arXiv.2104.10775

Gomathi, E., Jayasheela, M., Thamarai, M., & Geetha, M. (2023). Skin cancer detection using dual optimization based deep learning network. Biomedical Signal Processing and Control, 84, 104968. https://doi.org/10.1016/j.bspc.2023.104968

Sivakumar, N. R., Ghorashi, S. A., Karim, F. K., Alabdulkreem, E., & Al-Rasheed, A. (2022). MIoT Based Skin Cancer Detection Using Bregman Recurrent Deep Learning. Computers, Materials & Continua, 73(3), 6253–6269. https://doi.org/10.32604/cmc.2022.029266

Balambigai, S., Elavarasi, K., Abarna, M., Abinaya, R., & Arun Vignesh, N. (2022). Detection and optimization of skin cancer using deep learning. Journal of Physics: Conference Series, 2318(1), 012040. https://doi.org/10.1088/1742-6596/2318/1/012040

Fraiwan, M., & Faouri, E. (2022). On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning. Sensors, 22(13), 4963. https://doi.org/10.3390/s22134963

Pan, J.-S., Meng, Z., Li, J., & Virvou, M. (2022). Early Detection of Melanoma Skin Cancer Using Image Processing and Deep Learning. Springer. https://doi.org/10.1007/978-981-19-1053-1_25

Tembhurne, J. V., Hebbar, N., Patil, H. Y., & Diwan, T. (2023). Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-14697-3

Wang, X., Yang, Y., & Mandal, B. (2023). Automatic detection of skin cancer melanoma using transfer learning in deep network. AIP Conference Proceedings, 2562(1). https://doi.org/10.1063/5.0111909

Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing, 1–5. https://doi.org/10.1109/ATSIP49331.2020.9231544

Singh, H., Kaushik, S., Talyan, S., & Dwivedi, K. (2022). Skin Cancer Detection Using Deep Learning techniques. International Journal for Research in Applied Science and Engineering Technology, 10(5), 4296–4305. https://doi.org/10.22214/ijraset.2022.43090

Adebiyi, A., Abdalnabi, N., Smith, E. H., Hirner, J., Simoes, E. J., Becevic, M., & Rao, P. (2025). Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 and ISIC 2017 Datasets. medRxiv. https://doi.org/10.1101/2024.05.30.24308213

Shakya, M., Patel, R., & Joshi, S. (2025). A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification. Scientific Reports, 15, 4633. https://doi.org/10.1038/s41598-024-82241-w

Vieira, J., Mendonça, F., & Morgado-Dias, F. (2025). Deep Learning Approaches for Skin Lesion Detection. Electronics, 14(14), 2785.

Saeed, M. A., Afify, Y. M., Badr, N. L., et al. (2025). Multimodal deep learning ensemble framework for skin cancer detection. Scientific Reports, 15, 45660. https://doi.org/10.1038/s41598-025-30534-z

Yao, Y., Lilhore, U. K., Simaiya, S., et al. (2026). Advancing skin cancer diagnosis with deep learning and attention mechanisms. Scientific Reports, 16, 3610. https://doi.org/10.1038/s41598-025-33597-0

Skin Cancer Image Classification Using Artificial Intelligence Strategies: A Systematic Review. (2024). Diagnostics. Systematic review of 206 records on AI-based skin-tumor classification.

Yang, G., Luo, S., & Greer, P. (2025). Advancements in skin cancer classification: a review of machine learning techniques in clinical image analysis. Multimedia Tools and Applications, 84, 9837–9864. https://doi.org/10.1007/s11042-024-19298-2

Shaheen, H., & Singh, M. P. (2023). Skin lesion classification using HG-PSO and YOLOv7 based convolutional network in real time. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 237(10), 1228–1239. https://doi.org/10.1177/09544119231198823

Khurshid, M., Singh, R., & Vatsa, M. (2025). Multimodal dual-stage feature refinement for robust skin lesion classification. Scientific Reports, 15, 37775. https://doi.org/10.1038/s41598-025-14839-7

Li, X., Ouyang, J., Chen, Y., Diao, Z., Li, K., He, D., Liang, W., & Poniszewska-Marańda, A. (2026). Accurate skin lesion classification on imbalanced dermoscopic images with high variance via the SCTFD framework. Scientific Reports, 16, 12302. https://doi.org/10.1038/s41598-026-37846-8

Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, V. (2018). FiLM: Visual Reasoning with a General Conditioning Layer. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11671

Wilkes, M., Wright, C. Y., du Plessis, J. L., & Reeder, A. (2015). Fitzpatrick Skin Type, Individual Typology Angle, and Melanin Index in an African Population: Steps Toward Universally Applicable Skin Photosensitivity Assessments. JAMA Dermatology, 151(8), 902–903. https://doi.org/10.1001/jamadermatol.2015.0351