Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 Dataset

Abdulmateen Adebiyi MS[1], Nader Abdalnabi MS[2], Emily Hoffman Smith MD[5], Jesse Hirner MD[3], Eduardo J. Simoes MD[2,4], Mirna Becevic PhD[2,3], Praveen Rao PhD[1,2]

[1] Department of Electrical Engineering and Computer Science, University of Missouri, United States

[2] MU Institute for Data Science and Informatics, University of Missouri, United States

[3] Department of Dermatology, University of Missouri, United States

[4] Department of Biomedical Informatics, Biostatistics and Medical Epidemiology, University of Missouri, United States

[5]Department of Dermatology, Saint Louis University, United States

Corresponding Author: Praveen Rao, Ph.D., Department of Electrical Engineering and Computer Science, 411 S 6th St, Columbia, MO 65201

Abstract

Objectives: Our aim is to evaluate the performance of multimodal deep learning to classify skin lesions using both images and textual descriptions compared to learning only on images.

Materials and Methods: We used the HAM10000 dataset in our study containing 10,000 skin lesion images. We combined the images with patients' data (sex, age, and lesion location) for training and evaluating a multimodal deep learning classification model. The dataset was split into 70% for training the model, 20% for the validation set, and 10% for the testing set. We compared the multimodal model's performance to well-known deep learning models that only use images for classification.

Results: We used accuracy and area under the curve (AUC) receiver operating characteristic (ROC) as the metrics to compare the models' performance. Our multimodal model achieved the best accuracy (94.11%) and AUCROC (0.9426) compared to its competitors.

Conclusion: Our study showed that a multimodal deep learning model can outperform traditional deep learning models for skin lesion classification on the HAM10000 dataset. We believe our approach can enable primary care clinicians to screen for skin cancer in patients (residing in areas lacking access to expert dermatologists) with higher accuracy and reliability.

Keywords: Multimodal Deep Learning, Skin Lesion Classification, Primary Care Physician, Skin Cancer Screening

Lay Summary

Skin cancer, which includes basal cell carcinoma, squamous cell carcinoma, melanoma, and less frequent lesions, is the most frequent type of cancer. Around 9,500 people in the United States are diagnosed with skin cancer every day. Recently, multimodal learning has gained a lot of traction for classification tasks. Many of the previous works used only images for skin lesion classification. In this work, we used the images and patient metadata (sex, age, and lesion location) in HAM10000, a publicly available dataset, for multimodal deep learning to classify skin lesions. We used the model ALBEF (Align before Fuse) for multimodal deep learning. We compared the performance of ALBEF to well-known deep learning models that only use images (e.g., Inception-v3, DenseNet121, ResNet50). The ALBEF model outperformed all other models achieving an accuracy of 94.11% and an AUROC score of 0.9426 on HAM10000. We believe our model can enable primary care clinicians to accurately screen for skin cancer in patients.

Background and Significance

Skin cancer is the most common type of cancer diagnosed worldwide[1]. It is estimated that approximately 9,500 people in the United States (US) are diagnosed with skin cancer each year[2]. It is predicted that around 20% of people in the US will develop skin cancer[2]. The two most common skin cancer types are basal cell cancer and squamous cell cancer, while melanoma is the third most common skin cancer[2]. However, melanoma has the highest mortality, with a long-term survival of less than 10%, despite recent decline in mortality attributed to better treatment[3].

There are geographical differences in the incidence and mortality of skin cancer[4]. The incidence and mortality of melanoma is higher for individuals living in rural and underserved areas, than their urban counterparts[4]. Many factors may contribute to this geographical disparity in melanoma incidence and death, including increased ultraviolet radiation exposure and lower adoption of sun protection strategies in rural compared to urban residents[5]. In addition, barriers to health care access and availability contributes to late detection and effective treatment of skin cancers. The lack of adequate number and distribution of dermatologists contributes to late detection of melanoma. Patients face considerable wait times ranging from 33.9 to 73.4 days to consult with a dermatologist regarding changing moles. Interestingly, even when medical care is offered at no cost, a significant number of patients decline it if the travel distance for their appointment exceeds 20 miles[6].

This situation is accentuated in isolated rural areas. Tele-dermatology has proven effective in mitigating geographic isolation, and various studies have affirmed its diagnostic and treatment accuracy and reliability through telemedicine. Nonetheless, numerous obstacles hinder its widespread adoption and implementation. Alongside privacy and liability concerns, dermatologists identify the absence of a consistent reimbursement system as the primary impediment for both store-and-forward and live-interactive tele-dermatology[7]. Karavan et al.

discovered that 40% of patients diagnosed with melanoma in traditional clinics resided in areas where tele-dermatology services were underutilized[8].

Primary care clinicians based (PCCs) in the community often serve as the initial point of contact for patients and may assume a crucial role in offering screening and early diagnosis, particularly for individuals lacking sufficient access to dermatologists[7]. PCCs, while essential, face limitations compared to dermatologists in terms of early detection and have identified insufficient training during medical school and residency as hindrances to effective skin screening[9]. In primary care settings, the rapid and accurate diagnosis of skin conditions is of paramount importance for patient care[9]. However, distinguishing between various skin lesions can be a challenging task, and misdiagnosis can lead to serious consequences[10].

Prior work showed the effectiveness of state-of-the-art deep learning models for skin lesion classification (into malignant and benign classes) using dermascopy images[11]. With growing interest in multimodal deep learning models[12], it is now possible to combine skin lesion images and textual data (e.g., lesion location, patient age) for model training and inference. In this work, we investigate whether multimodal models can improve the accuracy of skin cancer diagnosis compared to models that only use images. Our approach, uses the Human Against Machine with 10000 training Images (HAM10000), a dataset comprising a diverse range of dermatological images[13] including basal cell cancer and melanoma.

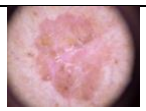
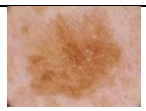
Methods

Dataset

In this study, we used the well-known HAM10000 dataset[13], which is a large dermatoscopic image collection of common pigmented skin lesions. The images were categorized into 7 classes, namely, Actinic Keratoses (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Melanocytic Nevi (NV), Melanoma (MEL), and Vascular Skin

Lesion (VASC). Table *1* shows examples of skin lesion images from each class along with the number of images per class.

Table 1: Examples of skin lesions in HAM10000

	AKIEC	**BCC**	**BKL**	**DF**	**NV**	**MEL**	**VASC**
Sample Image							
Count	327	514	1,099	115	6,705	1,113	142

In addition to the images, the dataset had 7 variables, which are described in Table *2*.

Table 2: Brief description of the variables in HAM10000

No.	**Variable**	**Brief Description**
1	Lesion id	Patient associated with skin lesion
2	Image id	Image name in the HAM10000 folder
3	Dx	Lesion class
4	Dx_type	How the skin lesion was verified (e.g., expert consensus, follow-up, using confocal microscopy)
5	Age	Age of patient
6	Sex	Sex of patient
7	Lesion Location	Location of the skin lesion (e.g., face, back, chest)

Deep Learning Models

In recent years, deep learning has gained a lot of traction in image understanding, image classification, language translation, speech recognition, and natural language processing. Convolutional neural networks (CNNs) have shown excellent performance in large-scale image classification and object detection competitions such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[14]. We present a few popular CNNs and the Multimodal that we used in our work.

Inception-V3: CNNs that are deep with many layers are prone to overfitting and consume a lot of computational resources. Inception-V3 was introduced by Szegedy et. al. in 2014[15] to solve these problems by using sparsely connected architectures. It uses the inception module that applies multiple convolutions (e.g., 1x1 convolution, 3x3 convolution, 5x5 convolution) and a maximum pooling layer. The outputs are concatenated to create the input for the next stage. A version of Inception-V3 called GoogLeNet with 22 layers won the ILSVRC 2014 competition[16]. Inception-V3 is an image recognition model that achieved around 78.1% accuracy on the ImageNet dataset. The model was first introduced and implemented in the paper "Rethinking the inception Architecture for Computer Vision"

ResNet: He et. al. introduced the deep residual neural network (ResNet) architecture, which won first place in the ILSVRC 2015 competition[17]. ResNet uses skip connections between layers to solve the vanishing gradient problem. Residual blocks reduced the total parameters by allowing the gradient to flow directly through the skipped connections backward from later layers to the initial filter. In our work, we use ResNet50 which has 50 layers.

DenseNet: Huang et. al. introduced the DenseNet[18] architecture where every layer in the model is connected to every other layer in a feed-forward manner. DenseNet was proposed to solve the problem of vanishing gradient while being computationally efficient. It promotes

feature reuse resulting in a more compact model. In our work, we use DenseNet121 which has 121 layers.

ALBEF: Junnan Li et al introduced ALBEF model[19].It is a state-of-the-art deep learning model that learns the joint representation of image and text data.The model combines the Vision Transformer (ViT-b/16) as the image encoder and BERT as the text encoder. We used the joint text-image encoder to encode both the text and images, and then add a linear fully connected layer to it. The Image encoder are initialized with weights pre-trained on ImageNet-1k[20]. The input image is encoded into a sequence of embeddings.

In our work, we used a joint text-image encoder which aligns the BERT text encoder's embeddings with the image encoder's (Vision Transformers). We then added a linearly fully connected layer and then predict the 7 output classes of AKIEC, BCC, BKL, DF, NV, MEL and VASC.

Data Pre-processing

Each original image in HAM10000 was of size 600×450 pixels. As different deep learning models used different image sizes as input, we had to resize the original images. For ResNet50 and DenseNet121, the images were resized to 224×224 pixels. For Inception-V3, the images were resized to 299×299 pixels. Finally, for ALBEF, the images were resized to 256×256 pixels. We also applied data augmentation techniques such as color jitter, random rotation, random horizontal flip, and random vertical flip to improve the performance of the trained models and prevent overfitting.

Skin Lesion Classification Approach

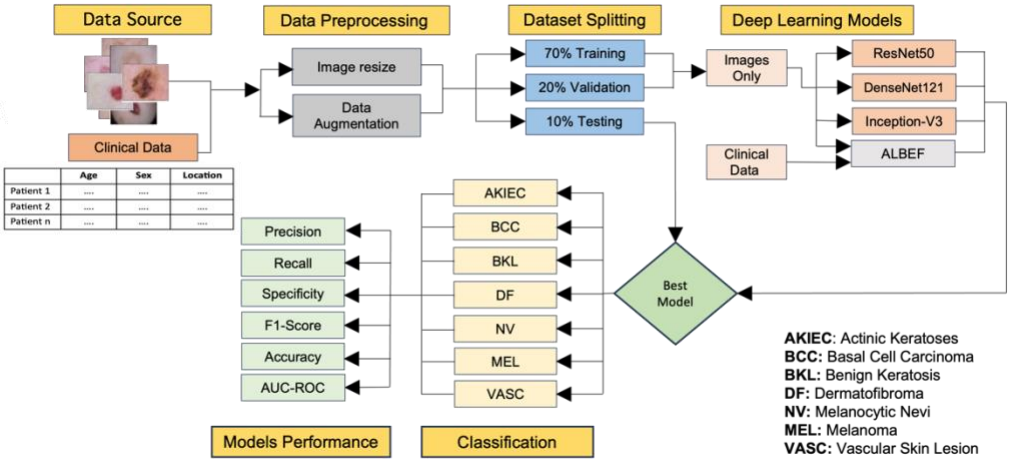


Figure 1: Skin lesion classification approach

We split the HAM10000 dataset into three sets randomly: the training set (containing 70% of the images), the validation set (containing 20% of the images), and the testing set (containing 10% of the images). The training and validation sets were used for the training phase. We then used the testing set to evaluate the trained models. The test set had the following number of images for each class: AKIEC: 38, BCC: 49, BKL: 110, DF: 11, MEL: 109, NV: 667, and VASC: 18.

Figure 1 shows the overall approach for skin lesion classification. Our system was implemented in Python using PyTorch[21], CUDA, Numpy[22], and OpenCV[23] libraries. We used existing implementations of the ALBEF model for our multimodal lesion classification[24]. The models were trained and tested on a Dell Precision server with an Intel Xeon processor,96 GB RAM, 2 TB disk storage, and two NVIDIA Quadro RTX4000 (8GB) graphics processing units (GPUs).

Model Training Settings

The Inception-V3, ResNet50, DenseNet121 and ALBEF models were trained with the same hyperparameters: (a) batch size of 16, (b) 200 epochs, and (c) learning rate of 1e-4. Also, each

model used the Adam optimizer[25] and binary cross entropy loss function. The best model based on the highest validation accuracy was saved and used for classification on the test set.

For the Inception-V3, ResNet50, and DenseNet121 model, we performed data augmentation on the lesions using RandomHorizontalFlip, RandomverticalFlip, RandomRotation, ColorJitter and we then transformed it to tensor using the ToTensor method in Pytorch. We trained 70% of the data using a batch size of 16 in 200 epochs. We then picked our best model by using 20% of the dataset as the validation set.

In our ALBEF model, we resized the image to 256 pixels. We then transformed the lesions using RandomResizedCrop and RandomHorizontalFlip. Our other hyperparameters are patch size of 16, Embedding dimension (embed_dim) of 768, depth of 12, num_heads of 12, weight_decay of 0.01, Multi-Layer Perceptron (mlp_ratio) of 4, Query Key Value (qkv_bias) of True and Epsilon (eps) of 1e-4. We trained 70% of the data using a batch size of 16 in 200 epochs. We then picked our best model by using 20% of the dataset as the validation set.

Results

In this section, we present the performance metrics of InceptionV3, ResNet50, DenseNet121, and the Multimodal fusion (ALBEF) on the HAM10000 dataset. On the ALBEF model, we performed two different settings.

1) Using the Images with the associated text (age, sex, and lesion location) for training the HAM10000 dataset.
2) Using only the Images and passing blank as the text for training on the HAM10000 dataset. We performed this experiment to show the effect of adding text on the overall performance of the model.

Performance Metrics

Next, we briefly describe the performance metrics that we used for evaluating our different models. The Classification models aim to classify the HAM10000 dataset into 7 classes. We used true positives (Tp), true negatives (Tn), false positives (Fp), and false negatives (Fn) in the computation of our different performance metrics. Below are the brief definitions of Tp, Fp, Fn, and Tn. Tp indicates the total number of HAM10000 Skin Lesions that were predicted correctly in the positive class, Tn denotes the lesions that are in the negative class and are classified correctly as the negative class. Fn denotes the total number of skin lesions that were in the positive class but were predicted incorrectly as the negative class and Fp are the lesions that are in the negative class but were predicted incorrectly as the positive class.

Table 3: Different metrics used in our evaluation

Metric	**Formula**
Precision (*P*)	$T_p/(T_p+F_p)$
Sensitivity/ Recall (*R*)	$T_p/(T_p+F_n)$
Specificity	$T_n/(T_n+F_p)$
Accuracy	$(T_p+T_n)/(T_p+T_n+F_p+F_n)$
F1-score	$2*P*R/(P+R)$

Table 4*: Comparison of the performance of the different models on the HAM10000 dataset (the best model is shown in* ***bold)***

Models	Accuracy	Precision	Sensitivity/Recall	Specificity	F1-Score	AUCROC
InceptionV3	0.8653	0.8202	0.7608	0.9570	0.7838	0.8589
ResNet50	0.8503	0.7815	0.7129	0.9646	0.7292	0.8388

DenseNet121	0.8862	0.8505	0.8249	0.9685	0.8350	0.8967
ALBEF (Only Images)	0.9132	0.8959	0.8524	0.9748	0.8693	0.9136
ALBEF (Text and Images)	**0.9411**	**0.9073**	**0.9019**	**0.9833**	**0.9033**	**0.9426**

Table 3 shows the performance metrics of InceptionV3, ResNet50, DenseNet121, and ALBEF on the HAM10000 dataset. ALBEF outperformed all other models. ALBEF model has the best accuracy, precision, recall, specificity, and F1-Score.

Discussion and Conclusion

We implemented the multimodal deep learning models to classify the skin lesions in the HAM10000[13] dataset. The sensitivity, specificity, AUCROC, accuracy and precision of classification using the ALBEF model with the image dataset and the associated text was the highest among the five different experiments that we conducted and deemed applicable in a primary care setting.

Multimodal model is a new field of Artificial Intelligence that helps to replicate the ability to combine information from multiple models. Information from different sources like audio, text, image, and video assists in implementing more complex models that improve the performance of many applications[26].Multimodal learning includes fusion-based approaches, alignment-based approaches, and late fusion.

The ALBEF model (Align Before Fuse) enables us to fuse visual and clinical information, providing a holistic view of the skin lesions[19].This Multimodal approach has the potential to significantly enhance the diagnostic capabilities of primary care physicians and nurse practitioners[27]. The ALBEF model used BERT as text encoder and Vision Transformer model as the image encoder. We used the joint text-image encoder to encode both the text and images and added a linear fully connected layer to it. ALBEF model has been used in a wide

array of domains such as Hateful detection, Image Retrieval etc[19]. ALBEF can learn joint representations which has made it very useful in Image retrieval tasks such as in text queries, text-based on images e-commerce applications, etc. It has also been applied in natural language processing for generating text captions for images. It can be used in classification tasks too.[24,28]

Adebiyi et al applied three different models (InceptionV3, ResNet50 and DenseNet121) for their skin lesion classification[11,29]. They collected 770 de-identified dermoscopy images from the University of Missouri (MU) Healthcare. They then created three unique images that contained the original images and images after they applied a hair removal algorithm. DenseNet121 achieved the best accuracy of 80.52% and an AUCROC score of 0.81 in their experiment.

Alam et al achieved an accuracy of 85% on the HAM10000 dataset by using the Inception-V3 model[30]. Akter et al achieved an accuracy of 0.82 on the HAM10000 dataset using the ResNet50 model. Our ALBEF model outperformed their works by achieving an accuracy of 0.9411[31].

Tschandi et al applied Multimodal learning for skin lesion classification[32].They employed two ResNet50 convolutional neural networks (CNN) followed by a late fusion technique to combine the features. Their result showed that combining the dermoscopic with macroscopic images and the metadata may improve network performance. Multimodal machine learning has been applied in Knee Osteoarthritis progression prediction from Plain Radiographs and Clinical Data by Tiulpin et al[33]. They utilized the raw radiographic data, clinical examination results, and previous medical history of the patient. They were able to achieve an area under the ROC curve (AUC) of 0.79.

We applied the Inception-V3, ResNet50, DenseNet121 and ALBEF models in our experiment. Our InceptionV3 model achieved an accuracy of 0.8653 and an AUCROC of

0.8589 on the HAM10000 dataset. The ResNet50 model achieved an accuracy of 0.8503 and an AUCROC of 0.8388. The DenseNet121 achieved an accuracy of 0.8862 and an AUCROC of 0.8967. We also implemented Multimodal models that combines the text features in the HAM10000 dataset and the Lesion for the Classification. The text we used in our experiments are the age, sex, and location of the lesion. The Multimodal model we used in this work is the well-known ALBEF (Align Before Fuse) model. We performed two experiments using the ALBEF model. In the first experiment, we used the lesion and passed the text as blank. This achieved an accuracy of 0.9132 and an AUCROC of 0.9136. In the second experiment on the ALBEF model, we used both image and the text. This achieved an accuracy of 0.9411 and an AUCROC of 0.9426. Overall, our Multimodal model outperformed all the other models. This shows that the addition of other patient's feature specifically text in this instance can improve the overall performance of the Skin Lesion Classification.

Our study has some limitations**.** We only used three metadata (Age, Sex, and Location) with the images in our multimodal classification. We believe having more metadata in the dataset may even increase our performance.

We recommend future studies utilize additional text features, as our study was limited to age, sex, and localization data.

Acknowledgments

This project was funded by the **T**ranslational **R**esearch **I**nforming **U**seful and **M**eaningful **P**recision **H**ealth (TRIUMPH) grant from the University of Missouri-Columbia.

Authors Contribution:

PR, EJS, MB, and EHS conceived the idea of multimodal learning on skin lesion images. AA implemented and evaluated the deep learning models. AA and PR designed the experiments and analyzed the results. JH and EHS provided clinical insights for the study. All authors were involved in drafting and editing the manuscript.

References

1. Working under the sun causes 1 in 3 deaths from non-melanoma skin cancer, say WHO and. https://www.iarc.who.int/cancer-type/skin-cancer
2. Skin cancer https://www.aad.org/media/stats-skin-cancer
3. Melanoma of the Skin - Cancer Stat Facts. Available from: https://seer.cancer.gov/statfacts/html/melan.html
4. Blake KD, Moss JL, Gaysynsky A, Srinivasan S, Croyle RT. Making the Case for Investment in Rural Cancer Control: An Analysis of Rural Cancer Incidence, Mortality, and Funding Trends. Cancer Epidemiol Biomarkers Prev. 2017 Jul;26(7):992–7.
5. Kalia S, Kwong Y k. k., Haiducu M l., Lui H. Comparison of sun protection behaviour among urban and rural health regions in Canada. Journal of the European Academy of Dermatology and Venereology. 2013;27(11):1452–4.
6. Pala P, Bergler-Czop BS, Gwiżdż JM. Teledermatology: idea, benefits and risks of modern age – a systematic review based on melanoma. Postepy Dermatol Alergol. 2020 Apr;37(2):159–67.

7. Jones OT, Jurascheck LC, van Melle MA, Hickman S, Burrows NP, Hall PN, et al. Dermoscopy for melanoma detection and triage in primary care: a systematic review. BMJ Open. 2019 Aug 20;9(8):e027529.
8. Karavan M, Compton N, Knezevich S, Raugi G, Kodama S, Taylor L, et al. Teledermatology in the diagnosis of melanoma. J Telemed Telecare. 2014 Jan;20(1):18–23.
9. Brown AE, Najmi M, Duke T, Grabell DA, Koshelev MV, Nelson KC. Skin Cancer Education Interventions for Primary Care Providers: A Scoping Review. J Gen Intern Med. 2022 Jul;37(9):2267–79.
10. Li H, Pan Y, Zhao J, Zhang L. Skin disease diagnosis with deep learning: A review. Neurocomputing. 2021 Nov 13;464:364–93.
11. Adebiyi A, Rao P, Hirner J, Anokhin A, Hoffman Smith E, Simoes E, and Becevic M. Comparison of Three Deep Learning Models in Accurate Classification of 770 Dermoscopy Skin Lesion Images. In AMIA 2024 Informatics Summit, 8 pages, Boston, 2024.
12. Huang Y, Du C, Xue Z, Chen X, Zhao H, Huang L. What Makes Multi-Modal Learning Better than Single (Provably). In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2021 [cited 2023 Dec 15]. p. 10944–56.
13. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018 Aug 14;5(1):180161.
14. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. arXiv; 2015

15. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. p. 1–9.
16. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2016 [cited 2023 Dec 15]. p. 2818–26.
17. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2016
18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
19. Li J, Selvaraju RR, Gotmare AD, Joty S, Xiong C, Hoi S. Align before Fuse: Vision and Language Representation Learning with Momentum Distillation [Internet]. arXiv; 2021
20. imagenet-1k · Datasets at Hugging Face 2024. https://huggingface.co/datasets/imagenet-1k
21. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2019
22. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357–62.
23. The OpenCV Library | https://opencv.org/
24. multimodal-learning-hands-on-tutorial/multimodal_training.ipynb at main · dsaidgovsg/multimodal-learning-hands-on-tutorial. https://github.com/dsaidgovsg/multimodal-learning-hands-on-tutorial/blob/main/multimodal_training.ipynb

25. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization [Internet]. arXiv; 2017
26. Pei X, Zuo K, Li Y, Pang Z. A Review of the Application of Multi-modal Deep Learning in Medicine: Bibliometrics and Future Directions. Int J Comput Intell Syst. 2023 Mar 29;16(1):44.
27. Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, et al. Multimodal machine learning in precision health: A scoping review. npj Digit Med. 2022 Nov 7;5(1):1–14.
28. ALBEF [Internet]. SERP AI. 2023 https://serp.ai/albef/
29. Adebiyi A, Flowers L, Giefer J, Hirner J, Rao P, Smith EH, et al. Accurate classification of benign and malignant dermoscopy skin lesions using three deep learning models. 2023
30. Alam TM, Shaukat K, Khan WA, Hameed IA, Almuqren LA, Raza MA, et al. An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset. Diagnostics. 2022 Sep;12(9):2115.
31. Akter MS, Shahriar H, Sweta Sneha. Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network. In: 2022 IEEE International Conference on Big Data (Big Data).
32. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. Exp Dermatol. 2018 Nov;27(11):1261–7.
33. Tiulpin A, Klein S, Bierma-Zeinstra SMA, Thevenot J, Rahtu E, Meurs J van, et al. Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. Sci Rep. 2019 Dec 27;9(1):20038.