COM6003 Artificial Intelligence and Machine Learning

Investigate the possibility of developing an Accurate Skin Cancer Detection System for Early Symptom Screening

Module instructor: Dr. Carlin CHU

Group ID: G4

Group member:

GE Yuxin（p233336）

KWOK Tsz Yi(P233340)

LEI Linlin（p233341）

WANG Yanyi（p233361）

Contents

Abstract

The incidence of skin cancer has been steadily increasing, necessitating innovative approaches to early detection and screening. In this report, we introduce a novel pre-screening tool that integrates computer vision and text analysis through deep learning and machine learning models to facilitate early detection of skin cancer. Our proposed system employs a two-pronged approach: firstly, a computer vision model analyses images of skin lesions taken by patients using mobile phone. Secondly, a separate machine learning model processes textual patient information, such as symptom descriptions and personal health history.

The computer vision model utilizes deep learning techniques, such as ConvMixer, ResNet-50, Visual Geometry Group (VGG), Big Transfer (BiT) and Segment Anything Model (SAM), to classify skin lesions based on visual patterns that correspond to benign or malignant characteristics. Parallelly, the text analysis model employs traditional machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), XG Boost (XGB), K-Nearest Neighbours (KNN) and Light Gradient-Boosting Machine (LGBM), to interpret relevant clinical information from patient-provided texts, which includes symptoms, duration, and previous medical conditions related to skin health.

The weighting and structure of these models will be ensembled via the stacking ensemble method, to ensure that the model performance is sustained. The integration of these models allows for a comprehensive analysis, improving the accuracy of preliminary diagnoses. This system is designed to be user-friendly, enabling patients to perform an initial check by simply taking a photograph of the skin lesion and inputting relevant textual information via a secure platform.

This report details the development process of our models, including data collection, model training, evaluating, tuning and deployment. Our findings suggest that this integrated approach can be a valuable tool in the early detection of skin cancer, potentially leading to timely medical intervention and improved patient outcomes.

Key words: Skin cancer, ConvMixer, ResNet-50, Visual Geometry Group, Big Transfer, Segment Anything Model, Stacking.

1. Introduction

1.1 Background and Objective

1.1.1 Background

In Hong Kong, the public healthcare system is facing significant challenges in managing the timely diagnosis of various diseases, including skin cancer. Currently, patients experience extended waiting times to consult with dermatology specialists in public hospitals, with some patients waiting for at least a year for an initial appointment. This delay is particularly concerning given that, according to the Hong Kong leading Cancer statistics 2020 [1], non-melanoma skin, a type of skin cancer ranks among the top 9 most common cancers in Hong Kong, however it takes at least 1 year for a new patient to have a consultation with a specialist in the Hong Kong public hospital. The prolonged waiting period can hinder early diagnosis and treatment, which are crucial for improving skin cancer prognosis.

Recent advancements in technology suggest a promising solution to this problem. Research in the field of computer vision has demonstrated potential in early skin cancer detection. Studies, such as Esteva et al [2]. in the journal Nature, have shown that deep learning algorithms can accurately classify skin cancer from images. This technological approach enables non-invasive, quick, and early detection of skin cancer, which could be particularly useful in regions with limited medical resources.

1.1.2 Objective

Given the critical situation in Hong Kong regarding the long waiting times for dermatological assessments and the rising incidence of skin cancer, our project aims to explore the opportunity of developing a computer vision-based system that enables earlier detection of skin cancer.

This project seeks to address the urgent need for reducing the burden on the healthcare system and improving patient outcomes through early detection. By allowing patients to take proactive steps in monitoring their skin health via simple photographic assessments and answering questionnaires, we hope to significantly shorten the diagnosis timeline and facilitate quicker medical interventions, hence potentially decreasing the mortality and morbidity associated with late-stage skin cancer diagnoses.

1.2 Data source

This study uses the Skin Cancer (PAD-UFES-20) dataset, which is derived from the skin lesion dataset composed of patient data and clinical images collected by smartphones by Andre et al. [3], and is the first A publicly available dermatology dataset. This data was launched on July 7, 2020 under the CC BY 4.0 DEED (Attribution 4.0 International), which allows the freedom to copy, redistribute and even use the material for commercial purposes in any medium or format.

This data set contains 2298 pictures and text data, including pictures of patient affected areas and basic patient diagnostic information.

1.3 Domain knowledge

Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Actinic Keratosis(ACK), Seborrheic Keratosis (SEK), Bowen's disease (BOD), Melanoma (MEL), and Nevus (NEV) are six types of skin diseases, comprising three dermatoses and three skin cancers, ranging from the more common basal cell carcinoma and squamous cell carcinoma to the rarer but more serious melanoma. In addition to basic demographic information, there are two types of text data features: Fitzpatrick skin type and biopsy [4].

Basal Cell Carcinoma (BCC): Basal cell carcinoma is a slowly growing skin cancer originating from cells in the basal layer of the skin. It typically presents as translucent or pearly nodules, sometimes with telangiectasia.

Squamous Cell Carcinoma (SCC): Squamous cell carcinoma is a malignant tumor originating from squamous cells in the skin. It usually presents as keratotic, proliferative, red, or pink nodules, sometimes with ulceration. SCC may metastasize to lymph nodes and other tissues.

Actinic Keratosis (AK): Actinic keratosis is a common skin lesion often caused by chronic exposure to ultraviolet radiation. It presents as rough, scaly, red, or brown patches on the skin surface, typically in sun-exposed areas.

Seborrheic Keratosis (SK): Seborrheic keratosis is a common benign skin lesion typically seen in older individuals. It manifests as raised plaques on the skin surface, varying in color from brown, black, to light yellow.

Bowen's Disease (BD): Bowen's disease is a precancerous lesion characterized by abnormal proliferation of squamous cells within the epidermis. It typically presents as red, rough patches or plaques, with a scaly surface. If left untreated, BD may progress to squamous cell carcinoma.

Melanoma (MEL): Melanoma is a malignant skin tumor originating from melanocytes in the skin. It usually presents as pigmented moles, varying in color from black to brown. Melanoma is characterized by its potential for rapid growth and metastasis to distant sites.

Fitzpatrick skin: The Fitzpatrick skin type is a commonly used system to describe individuals' tolerance to sunlight and their skin's response to ultraviolet radiation. This system typically categorizes individuals into six types based on their skin color, history of sunburns, and skin reactions.

Biopsy: "Biopsy" refers to the process of obtaining a small sample of tissue or cells from the human or animal body through surgery or other medical procedures for further examination, diagnosis, or research purposes. This procedure is typically performed by physicians or pathologists and can be used to identify abnormal cells, lesions, or diseases within the tissue, aiding in the formulation of appropriate treatment plans.

2. Data processing

2.1 Data Preprocessing

This project involves two parts of data: text and pictures, and these two types of data are preprocessed respectively.

2.1.1 Text data processing

Among the 26 features of the text data, consider that 13 columns of data have missing values and are filled with the constant value '-1'. Since the purpose of this project is to classify the types of skin accidents and thereby improve the recognition accuracy, the meaningless columns 'patent_id', 'lesion_id', and 'img_id' are deleted regardless of the relationship between the patient and the lesion.

At the same time, considering that there is a strong correlation between the 'biopsied' column and the result column, that is, all patients with skin cancer achieved complete spasticity, to avoid overfitting, the 'biopsied' column was deleted.

In response to the data imbalance problem, in order to ensure that the data can better learn the characteristics of the minority classes during the training process, while avoiding the excessive influence of the majority of classes on the model, we use a combination of up sampling and under sampling to deal with it [5].

2.1.2 Image data processing

Since the pixels of the image data in the original data set are not uniform, we resized and standardized the image data to make it a three-channel image of 256*256 pixels. To address the problem of image data imbalance, we use the following method to process image data.

2.2 Feature engineering

2.2.1 Feature engineering of text data

In order to make the model training effect better, we also extracted features from the data, used SelectKBest for feature selection, and used cross-validation to evaluate the performance of different feature subsets, thereby selecting the best feature subset of 9 columns, but because by comparing the training results of the three methods of best subset, deleting empty columns, and retaining all columns, it was found that the best results all performed best in the data set retaining all columns, so we finally used all Column for model training. At the same time, we performed Ordinal Encoding on all categorical type columns and used min-max scaler for normalization.

2.2.2 Feature engineering of image data

In order to improve the generalization ability of the image model during training and enable the model to learn image features under different conditions, we enhanced the image, including random left and right transformation, random up and down transformation, random rotation, random noise adjustment and random recovery adjustment.

2.3 Exploratory data analysis on AWS Platform

Using AWS to execute Python code can improve development efficiency and application flexibility, save costs, and provide scalable solutions while ensuring security and global access capabilities. Our data exploration and analysis of the project was performed on the AWS platform.

We identified several factors associated with potential risk of skin cancer. Specifically, people of German or Pomeranian descent, those aged 41 to 80, men, and people with long-term exposure to pesticides or chemicals, as well as people living in areas lacking clean water or sewage treatment facilities, appear to be more susceptible. Vulnerable to skin cancer. We also found that individuals whose skin was also more tolerant to the sun had a lower risk of skin cancer.

In terms of correlation of dermatological symptoms, BCC usually presents with lumps and itching of the incisions, and SCC has mild symptoms but also includes itching and swelling of the incisions. MEL grows rapidly and for ACK, NEV and SEK. ACK is mostly related to epidermal itching, while NEV and SEK have lumps as the main symptom. These findings provide important clues for identifying high-risk groups and symptoms, helping to develop early diagnosis and treatment strategies.

3. Model training and tuning

3.1 Model definition

This project mainly uses the stacked ensemble learning model [6] to perform integrated learning on five text data models and five image data models. The integrated learning results are compared with the meta-model, and finally the models with better results are selected. Form a comprehensive ensemble model to collaboratively analyze and predict the classification of skin lesion types.

The models based on image data principally utilize Very Deep Convolution (VGG), ConvMixer, ResNet50, SAM, and Big Transfer (BiT) for image classification analysis.

VGG [7] model is a relatively simple and traditional convolutional neural network composed of convolutional and pooling layers without complex modules. The VGG model is a straightforward, traditional convolutional neural network architecture that primarily utilizes sequential layers of convolutions and max pooling, without incorporating complex modules like residual connections or attention mechanisms. It's characterized by its deep structure, typically 16 to 19 layers, and employs 3x3 convolution filters and 2x2 pooling filters systematically to process visual data efficiently.

ResNet50 [8], a deep residual network, incorporates skip connections in its 50-layer structure to address the vanishing gradient problem prevalent in deep neural networks. The design of each residual block ensures lossless information transfer across deep network layers, significantly enhancing training efficiency and predictive accuracy. In this project, ResNet50 is employed to capture deep image features and enhance the model's ability to recognize characteristic disease markers.

The ConvMixer mode[l] [9], a relatively innovative neural network, combines features of traditional convolutional networks and mixture of experts systems to efficiently process large-scale image data. The ConvMixer model consists of a convolutional layer, multiple ConvMixer blocks, and a global average pooling layer. Each ConvMixer block includes depthwise separable convolutions, batch normalization, GELU activation functions, and residual connections. The output of the ConvMixer model is processed through global average pooling, followed by a fully connected layer to produce the final classification results.

The SAM model [10] is specifically designed for image segmentation, utilizing a flexible architecture to precisely identify and segment various objects within images. This capability is particularly crucial for extracting key information such as lesion areas in skin disease image analysis, thereby enhancing the project's segmentation accuracy and providing refined image features for disease classification.

The BiT model [11], pre-trained on a large-scale dataset, learns rich visual feature representations, optimizing the model's generalization ability and adaptability to various visual tasks. The multi-task learning strategy of BiT enables effective utilization of knowledge acquired during pre-training when transferred to specific medical image analysis tasks, thus improving the performance and stability of the entire ensemble system.

In the textual data, models such as Support Vector Machine (SVM), Random Forest (RF), XGBoost (XGB), K-Nearest Neighbors (KNN), and Light Gradient Boosting Model (LGBM) are primarily used for classifying whether skin cancer is present.

LGBM [12] is an efficient gradient boosting model based on a histogram algorithm that compresses data to reduce memory usage. Additionally, it employs a leaf-wise strategy to minimize loss, thus achieving better accuracy.

Most importantly, this project uses stacked ensemble learning. Stacking improves prediction accuracy by combining the predictions of multiple base learners to train a new meta-model.

During this process, various base models such as RF and SVM are independently trained and used to make predictions. These predictions serve as meta-features and can be used together with the original features or independently to form a new data set for the second-layer meta-model to learn. The purpose of a metamodel is to learn how to most efficiently combine inputs from various base models to optimize the final output. This approach leverages the diversity and strengths of different models to enhance generality and, through effective integration of meta-models, often provides performance beyond that of any single model, making it particularly suitable for complex classification or regression problems.

3.2Model Training

3.2.1 Training

The model training in this project is executed in two distinct phases. The initial phase entails training individual models independently and generating their predictions. The subsequent phase involves the integration of meta-features through a stacking ensemble method, leading to the training of a meta-model.

During the first phase, models are trained separately according to the data type—textual or image. For text-based skin disease classification, the data is first divided into training and testing sets, with a further 20% of the training set reserved as a validation set to prevent overfitting. Traditional classifiers like SVM, RF, and KNN are initially employed. However, their suboptimal performance on the test set prompted the inclusion of more sophisticated models such as XGBoost and LGBM.

For image data, the dataset undergoes a similar partitioning into training, validation, and testing sets. A suite of five models—VGG, ConvMixer, ResNet50, SAM, and Big Transfer—are meticulously selected and calibrated to exploit their classification strengths. Notably, the Big Transfer (BiT) model from TensorFlow Hub is fine-tuned by zeroing the kernel_initializer to diminish the impact of pre-trained weights and replacing the original classification layer with a Dense layer matched to the number of desired classes.

In the second phase, separate stacking ensembles for text and image data are constructed. Here, base models yield meta-features to train meta-models that harness predictions from their respective data type, enhancing predictive precision.

Ultimately, the efficacy of the two meta-models—one for text data and the other for image data—is evaluated. The model demonstrating superior accuracy and generalization on novel data is chosen for the final deployment. This methodical selection capitalizes on the specific strengths of each data-centric model, thereby bolstering the predictive robustness of the framework.

3.2.2 Preliminary Model Evaluation

Based on text data processing, the LGBM model outperformed others, achieving a precision of 0.63, a recall rate of 0.43, and an accuracy of 0.43. However, when a stacking ensemble method was applied to combine multiple models, the accuracy slightly decreased to 0.41, indicating that the inclusion of lower-performing models may have diluted the overall effectiveness.

For image data analysis, the SAM model was the frontrunner with an accuracy of 0.38, with ConvMixer trailing at 0.30. In stark contrast, models like Big Transfer, Resnet50, and VGG lagged significantly, with accuracies not exceeding 0.01.

The above findings suggest that the current model configurations and training strategies may not be adequately suited for image data, and there may be a need to explore improved model architectures and training parameters to enhance overall model performance and reliability.

Text Model:

Model	Precision	Recall	Accuracy
SVM	0.54	0.38	0.40
RF	0.59	0.43	0.43
XGB	0.62	0.42	0.42
KNN	0.52	0.39	0.40
LGBM	0.63	0.43	0.43
Stacking	0.59	0.40	0.41

Image Model:

Model	Precision	Recall	Accuracy
Resnet 50	0.01	0.01	0.01
VGG	0.01	0.01	0.01
Big Transfer	0.11	0.10	0.10
ConvMixer	0.09	0.30	0.30

SAM	0.15	0.38	0.38
Stacking	0.31	0.19	0.19

3.3 Model tuning

Based on the evaluation results presented in the previous section, the text analysis model performed way better than the image analysis model; the stacking model of text analysis has an accuracy of 0.41, with precision of 0.59 and recall of 0.40, while the stacking model of image analysis has an unsatisfactory performance of accuracy 0.19, having 0.19 recall and 0.31 precision. Although there is no compulsory requirement from HKSAR department of Health, as highlighted by Habehh and Gohel [13], the general accuracy of a machine learning model in the healthcare sector, especially in the field of diagnosis, has a minimum of 81.30% accuracy for preliminary public launchable. A pre-screening model having only a maximum of 61.6% accuracy for skin cancer diagnosis, is insufficient for public usage.

To enhance the accuracy of our model, we chose to simplify the output classification from six types — BCC, MEL, SCC, ACK, NEV and SEK—to two categories: skin cancers (BCC, MEL, SCC) and skin diseases (ACK, NEV, SEK). This simplification addressed the issue of data imbalances, as the proportions of skin cancer (48%) and skin diseases (52%) are now similar. Consequently, the balanced data distribution eliminated the need for under sampling, which previously led to a substantial data loss of 75%. We are now able to utilize the entire dataset for training our model, optimizing its learning potential.

Typically, a SoftMax function is employed for multi-class classification, which calculates a probability distribution across all classes, causing interdependent class probabilities. The denominator $(\sum_{j=1}^{K} e^{z_j})$ of the SoftMax function acts as a normalization constant, ensuring that sum of all probabilities from $\sigma(z)$ equals to 1, if there are too many variables classified as a particular class type, the probability of other variables being classified in that class type will be reduced. In contrast, the sigmoid function is more appropriate treats each output independently.

Sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

SoftMax function:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{Z_j}} \; for\ i = 1, \dots, K\ and\ z = (z_1, \dots, z_K)$$

Where:

- z is the score from the classifier
- K is the total number of classes

Given that the probability of each patient having skin cancer is assumed to be independent, utilizing the sigmoid function is likely more suitable than the SoftMax function, which might inaccurately couple the probabilities of unrelated outcomes.

3.4 Model performance comparison

The performance of the models developed for text and image analysis in our skin cancer detection system reveals insightful outcomes, particularly when considering the complexity of categorizing skin lesions into six classes versus a simpler binary classification.

For the text model, the LGBM exhibits the highest precision (0.63 for six classes and 0.93 for two classes), recall (0.43 for six classes and 0.93 for two classes), and accuracy (0.43 for six classes and 0.93 for two classes), making it the most effective among the tested models. RF and XGB models also show strong performance, particularly in binary classification with accuracies of 0.92. The stacking approach, which combines multiple models, presents a balanced performance with accuracies of 0.41 for six classes and 0.91 for two classes.

In the image model analysis, for six classes classification the SAM model perform the best, having 0.38 of accuracy, 0.38 recall and 0.15 precision. For binary class classification, VGG model perform the best, having 0.72 of accuracy, 0.74 precision and 0.72 recall. The Stacking model, which integrates outputs from multiple architectures, also performs commendably with accuracies of 0.19, precision of 0.31 and recall of 0.19 for six classes; accuracies of 0.52, precision of 0.52 and recall of 0.52 for binary classes.

Text Model:

Model	Precision		Recall		Accuracy	
No. of classes	Six	Two	Six	Two	Six	Two
SVM	0.54	0.87	0.38	0.87	0.40	0.87
RF	0.59	0.93	0.43	0.92	0.43	0.92
XGB	0.62	0.92	0.42	0.92	0.42	0.92
KNN	0.52	0.86	0.39	0.86	0.40	0.86
LGBM	0.63	0.93	0.43	0.93	0.43	0.93
Stacking	0.59	0.91	0.40	0.91	0.41	0.91

Image Model:

Model	Precision		Recall		Accuracy	
No. of classes	Six	Two	Six	Two	Six	Two
ConvMixer	0.09	0.22	0.30	0.47	0.30	0.47
Resnet 50	0.01	0.22	0.01	0.47	0.01	0.47
VGG	0.01	0.74	0.01	0.72	0.01	0.72

BiT	0.11	0.74	0.10	0.58	0.10	0.58
SAM	0.15	0.55	0.38	0.54	0.38	0.54
Stacking	0.31	0.52	0.19	0.52	0.19	0.52

For text analysis, we have opted to use a stacking model for meta-stacking. This decision is based on the relatively narrow performance range among the models we evaluated. The best model is only 0.1 percentage points more accurate than the least effective model. Employing a stacking approach helps ensure that the final model maintains consistent performance stability, capitalizing on the strengths of each individual model without significant trade-offs.

For image analysis, however, we adopt a different strategy by choosing the best-performing model for meta-stacking. In this case, the disparity in performance among the models is much more pronounced, with the top model outperforming the lowest by a substantial margin of 0.5 in accuracy. Directly using the best model for stacking in image analysis ensures that we leverage its superior performance while avoiding the potential dilution of effectiveness that might occur if less accurate models were included in the stack.

Hence, in the stage of meta-model diffusion, for six classes classification we choose to stack the text stacking model with image SAM model; for binary classification we choose to stack the text stacking model with image VGG model. These strategies are tailored to address the specific challenges and performance characteristics of text and image analysis within our system, ensuring that each approach is optimized for the best possible outcome.

Stacked model performance for Image and Text:

Model	Precision		Recall		Accuracy	
No. of classes	Six	Two	Six	Two	Six	Two
Stacking	0.37	0.83	0.39	0.82	0.39	0.82

4. Conclusion

4.1 Deployment

In order to deploy the model with an user interface, we can adopt Amazon Web Services (AWS). Initially, the application's backend, which handles data intake and processing, is hosted on Amazon EC2 instances, providing a secure and scalable environment. The patient questionnaire and image upload interface are integrated into a web service hosted on AWS Elastic Beanstalk, which automates scaling and management of the application layers.

Data from the questionnaire and images are stored securely on Amazon S3, ensuring high availability and durability. The text and image classification models are developed and trained using Amazon

SageMaker, a platform that facilitates the creation, training, and deployment of machine learning models at scale.

Once the models are trained, they are deployed as SageMaker endpoints. The application uses these endpoints to send the user-input data for real-time analysis. The text and image data are processed independently by their respective models, with each generating preliminary outputs.

The stacking technique is implemented using AWS Lambda functions, which are triggered once both models have processed the data. These functions handle the logic of combining the outputs from the text and image models and computing the final binary classification.

Once a user interacts with the system, they complete a digital questionnaire and upload necessary images via the web service. The result, indicating '0' for no cancer or '1' for cancer, is then returned to the user through the web interface.

4.2 Conclusion

Our project explores the feasibility of enabling the public to conduct preliminary skin cancer screenings independently, and preliminary results confirm its viability. As the result, the meta model of traditional machine learning model on patient textual information and deep learning model on patient skin image under binary classification output has a satisfactory performance, i.e. 0.82.

While the standalone text model achieves a high accuracy of 0.91, integrating it with the best-performing image model results in a reduced accuracy of 0.82. Despite this, it is advisable to combine the use of questionnaires and imaging for a more comprehensive analysis. This approach not only broadens the diagnostic scope by incorporating both structured questionnaire data and visual imaging but also reduces bias in questionnaire responses, where users might not be fully aware of their health conditions, leading to potentially misleading data.

Conversely, image analysis offers objective insights that are not influenced by user interpretation, providing a more reliable basis for diagnosis. By integrating both data sources, the system leverages the strengths of each method, complementing the potential bias from questionnaires with the objective evidence from imaging, thus enhancing the overall diagnostic accuracy and reliability of our skin cancer detection system.

However, before developing this into a publicly available skin cancer pre-screening application, several critical factors must be addressed. These include assessing potential biases within our model and ensuring the credibility of the data used for model training.

References

[1] H. A. Hong Kong Cancer Registry, "Leading Cancer Sites in Hong Kong in 2020," Hong Kong Cancer Registry, Hospital Authority, Hong Kong, 2020.

[2] E. e. al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature (London),* vol. vol. 542, no. no. 7639, p. pp. 115–118, 2017.

[3] L. G. S. A. K. B. B. I. d. A. G. A. F. J. E. J. S. A. C. P. R. F. F. P. K. R. K. H. S. M. d. E. S. R. M. T. C. T. d. B. L. Pacheco AGC, "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data Brief.," 2020 Aug.

[4] A. C. Society, "What Is Basal and Squamous Cell Skin Cancer?," [Online]. Available: https://www.cancer.org/cancer/types/basal-and-squamous-cell-skin-cancer/about/what-is-basal-and-squamous-cell.html.. [Accessed 15 Apr 2024].

[5] S. H. a. F. K. F. Thabtah, "Data imbalance in classification: Experimental evaluation," *Information Sciences,* vol. 513, pp. 429-441, 2020.

[6] M. e. a. Mohammed, "A stacking ensemble deep learning approach to cancer type classification based on TCGA data," *Scientific Reports,* vol. 11, no. 1, p. 15626, 2021.

[7] B. Ummepure, "Classification of Skin Cancer Using ResNet and VGG Deep Learning Network," *in Proceedings of the International Journal of Computer Applications,* vol. 184, no. 42, 2023.

[8] D. H. a. Z. Jie, "Research on Skin Cancer Classification Method Based on Improved ResNet50," *Computer Technology and Development,* vol. 2023, no. 02.

[9] A. T. a. J. Z. Kolter, "Patches Are All You Need?," *arXiv:2201.09792.* .

[10] Y. L. a. X. Y. M. Hu, "SkinSAM: Empowering Skin Cancer Segmentation with Segment Anything Model," *arXiv:2304.13973.*

[11] A. K. e. al., "Big Transfer (BiT): General Visual Representation Learning," *ar5iv.org,* 2023.

[12] Q. M. T. F. T. W. W. C. W. M. Q. Y. a. T.-Y. L. G. Ke, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Advances in neural information processing systems,* no. 30, 2017.

[13] H. a. S. Gohel, "Machine Learning in Healthcare," *Current genomics,* vol. 22, no. 4, p. 291–300, 2021.