Challenge Report: Dermoscopic Image Classification

Télécom Paris
Institut Polytechnique de Paris
Laura Manuela Castañeda Medina
castaneda@telecom-paris.fr

I. Introduction

Skin cancer, particularly melanoma, is one of the most lethal forms of cancer due to its high potential for metastasis. Despite being treatable through early detection and surgical excision, the incidence of melanoma and other skin cancers has risen significantly over recent decades, especially in populations with lighter skin tones. According to the World Health Organization [5], between 2 and 3 million non-melanoma skin cancers and 132,000 melanoma skin cancers occur globally each year. Early detection is critical as it significantly enhances treatment outcomes and reduces mortality.

For this reason, the field of dermatology has increasingly turned to non-invasive techniques for early diagnosis. Among these, computer-aided diagnosis systems have shown promising potential to assist dermatologists by enhancing the accuracy and efficiency of skin lesion analysis. These systems leverage machine learning algorithms to classify skin lesions based on dermoscopic images, potentially improving diagnostic accuracy and accessibility of skin cancer screening.

The goal of the proposed Kaggle challenge was to develop a robust system capable of classifying dermoscopic images into one of eight diagnostic categories: Melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis, Benign keratosis, Dermatofibroma, Vascular lesion, and Squamous cell carcinoma. This classification is based on visual and textural differences of the lesions, guided by the ABCD rule (Asymmetry, Border irregularity, Colour, Dimension), which is a well-established method in dermatology for assessing the potential malignancy of pigmented skin lesions.

This report will detail the methodologies employed in feature extraction, model selection and evaluation.

II. Dataset Description

Participants of the challenge were provided with a dataset consisting of 25,331 dermoscopic images, accompanied by metadata such as age, sex, and anatomical position of the lesions. The dataset is divided into a training-validation set and a test set, with classifications provided only for the training set. This structure challenge participants to develop and train machine learning models that can generalize well to unseen data, using only the features extracted from the images and metadata.

A. Data Composition

- **Images**: Each entry in the dataset represents a dermoscopic image of a skin lesion. The images vary in terms of size, color, and complexity, reflecting the diverse nature of skin lesions.

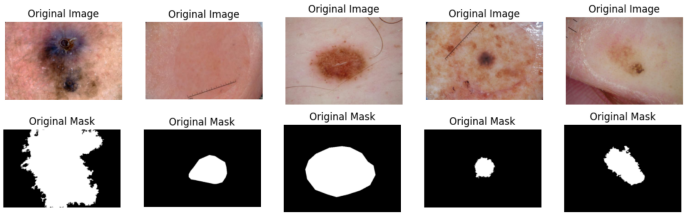

Figure 1. Original Images and Masks

- **Metadata**: Accompanying each image is metadata that includes the patient's age, sex, and the anatomical position of the lesion. This metadata is essential for the feature extraction phase, as certain skin cancers are more prevalent in specific demographics or body locations.

ID	CLASS	SEX	AGE	POSITION
ISIC_0028766	2	male	30.0	NaN
ISIC_0071222	8	male	85.0	lower extremity
ISIC_0069434	3	male	85.0	head/neck
ISIC_0062098	1	male	55.0	head/neck
ISIC_0057224	8	female	45.0	lower extremity

Table I
Initial metadata provided for the challenge

B. Classes

The images are categorized into eight diagnostic classes, each corresponding to a distinct type of skin lesion:

1) Melanoma
2) Melanocytic nevus
3) Basal cell carcinoma

4) Actinic keratosis
5) Benign keratosis
6) Dermatofibroma
7) Vascular lesion
8) Squamous cell carcinoma

These classes cover a range of benign and malignant skin conditions, from common moles (Melanocytic nevus) to more severe cases such as Melanoma and Basal cell carcinoma.

C. Data Split

The dataset is pre-divided into two subsets:

- Training-Validation Set (75%): This set includes both the images and their corresponding class labels. It is used for training the machine learning models and validating their performance during the development phase.
- Test Set (25%): This set consists of images without class labels. The performance of the models is evaluated on this subset, and the results are used to rank participants in the challenge.

D. Purpose and Use

The primary purpose of this dataset is to enable participants to develop and refine machine learning models that can accurately classify types of skin lesions based on dermoscopic images. By providing a diverse and comprehensive set of images along with relevant clinical metadata, the challenge aims to mirror real-world conditions where a CAD system would be employed, thus ensuring the practical applicability of the solutions developed.

III. Data Preprocessing

A. Imbalance Analysis

A preliminary analysis of the dataset revealed significant class imbalances, which is a common challenge in medical image classification tasks. Class imbalances can severely affect the performance of machine learning models, leading to biases towards more frequently represented classes.

The training set contained disparities in the number of images per class, with Class 2 (Melanocytic nevus) having over 9,000 images, while Classes 4 (Actinic keratosis), 6 (Dermatofibroma), 7 (Vascular lesion), and 8 (Squamous cell carcinoma) each had fewer than 500 images. Such discrepancies pose a risk of model overfitting to the more common classes and underperforming in accurately identifying rarer conditions.

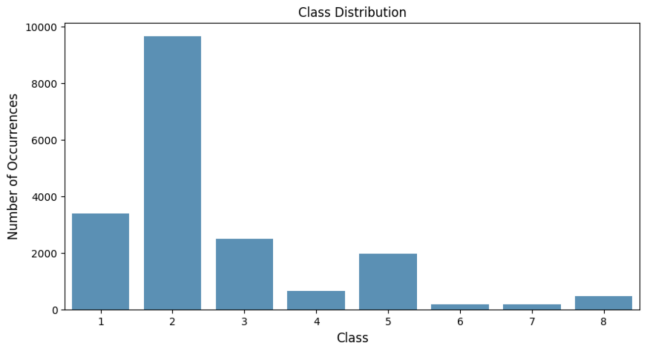

Figure 2. Original Class Distribution

The skewed distribution highlighted the necessity for strategic interventions to ensure that the developed models would generalize well across all types of skin lesions and not just the ones most represented in the dataset.

B. Data Augmentation

To mitigate the class imbalance identified in the dataset, data augmentation techniques were employed using the `ImageDataGenerator` class from TensorFlow's Keras preprocessing library. This approach helps in artificially expanding the size of a dataset by generating transformed versions of images, thereby providing more varied examples for model training. This can significantly improve the robustness and generalization ability of the model. Here is how the augmentation was implemented:

Techniques Used:

- *Rotation*: Images were randomly rotated by up to 20 degrees to simulate different orientations of skin lesions.
- *Width and Height Shifts*: Images were translated horizontally and vertically by 10% of their width and height, respectively, to simulate variation in lesion positioning.
- *Shear Transformation*: A shear intensity of 0.1 was applied, which distorts the image along an axis, typically creating a slant. This simulates changes in the angle of imaging.
- *Zoom*: Random zooming of images by up to 10% helps the model learn to recognize features at different scales.
- *Flips*: Horizontal and vertical flips simulate mirror images of lesions, addressing the orientation variance.
- *Fill Mode*: The 'nearest' fill mode was used for filling in newly created pixels, which helps maintain the integrity

of the lesion's texture.

This setup was used to augment the images of the underrepresented classes, effectively balancing the dataset. The augmented data ensures that each class had approximately 7,000 images, except for Class 2, which retained a slightly higher count due to its initial abundance. This strategy aimed to not only balance the number of images across classes but also to introduce sufficient variability, enhancing the model's ability to generalize across unseen data.

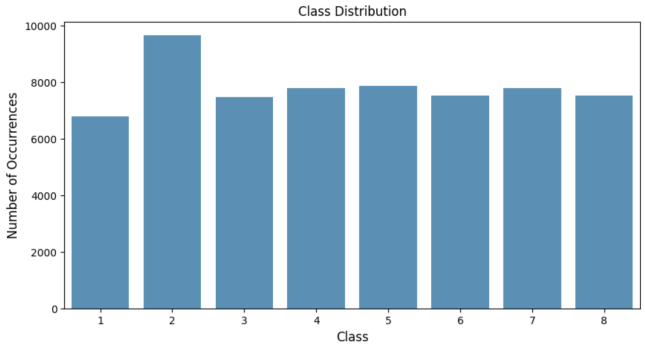

Figure 3. Balanced Class Distribution

C. Contrast Enhancement

To further enhance the visibility and distinguishability of skin lesions in the augmented dataset, contrast stretching was implemented. This preprocessing technique is crucial as it improves the image quality by increasing the dynamic range of intensities in the image. Enhanced contrast allows for better feature extraction, as more details become noticeable, which is particularly beneficial for medical image analysis.

- **Color Space Conversion:** The images were converted from the BGR color space to YUV. This conversion is instrumental because the YUV color space separates the luminance (Y) from the chrominance (U and V) components. By isolating the luminance, histogram equalization can be applied specifically to the brightness without altering the color information.

- **Histogram Equalization:** This technique improves the contrast in images by effectively spreading out the most frequent intensity values, thus enhancing areas of lower local contrast.

- **Reconversion:** After adjusting the luminance, the images are converted back to the BGR color space, which is more suitable for displaying and further processing in most computer vision tasks.

This function was applied to each image in the dataset after the augmentation process. By enhancing the contrast, we ensure that subsequent steps in feature extraction and model training can leverage the improved quality and visibility of key lesion characteristics.

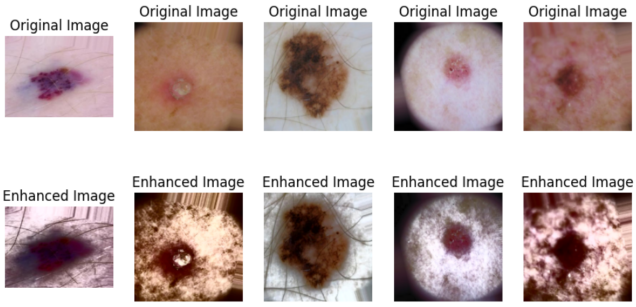

Figure 4. Enhanced Images

IV. Segmentation

Segmentation is a fundamental process in the analysis of medical images, as it allows for the precise identification and isolation of regions of interest— in this case, skin lesions—from the rest of the image. Effective segmentation is crucial for accurate feature extraction, particularly when applying the ABCD rule in later stages of the analysis.

A. Existing Masks

The dataset provided included pre-existing segmentation masks for approximately 2,000 images. These masks are crucial as they offer a ground truth for the lesion boundaries, allowing for more accurate feature calculations. In the presented workflow, these pre-existing masks were directly used where available, ensuring consistency in the initial set of segmented data.

B. Mask Generation

For images without pre-existing masks, a methodical approach was employed to generate new segmentation masks. The steps involved are designed to handle the inherent variability in skin lesion images, such as differences in color, size, and border irregularity. The following methods were used:

1) **Gaussian Blur:** Initially, each image was subjected to Gaussian blurring. This step reduces image noise and detail, which helps in achieving a more homogeneous region that enhances the effectiveness of thresholding.

2) **Otsu's Thresholding:** After blurring, Otsu's method was applied to perform global image thresholding. This technique automatically determines the optimal threshold value for converting a grayscale image to binary (black and white), which is essential for distinguishing the lesion from the background.

3) **Mask Inversion:** Since Otsu's thresholding may sometimes highlight the background instead of the lesion, an inversion step was incorporated. This ensures that the lesions are white (indicating the area of interest) and the background is black.

Generated masks were saved in a separate folder and linked to their corresponding images. This organization facilitates ease of access and reference during the feature extraction phase. Ensuring accurate segmentation across all images, whether through existing or newly generated masks, is crucial for the reliability and accuracy of subsequent analyses.

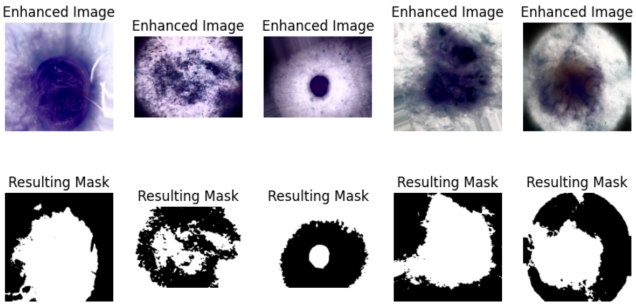

Figure 5. Generated Masks

V. Feature Extraction

The ABCD rule provides a systematic method for assessing skin lesions and is widely used in dermatology for the early detection of melanoma. Each component (Area, Border, Color, and Dimension) captures different aspects of the lesion, which can indicate benign or malignant characteristics.

A. Geometric Features [4]

- **Area and Perimeter:** The area of a lesion and the perimeter give fundamental clues about lesion growth. Malignant lesions, such as melanomas, often show rapid changes in size compared to benign lesions like nevi. Monitoring changes in these metrics can be critical for early cancer detection.

- **Circularity:** Circularity measures how close the lesion's shape is to a circle. Benign lesions are more likely to be circular, whereas malignant lesions often have irregular, non-circular shapes due to uneven growth rates.

- **Bulkiness:** Represents how voluminous a lesion appears relative to its boundary. This feature can help in distinguishing raised lesions, which are more common in benign cases like dermatofibromas.

- **Solidity:** This measures the proportion of the lesion's area compared to its convex hull. Lower solidity can indicate irregular or jagged borders, a common trait in malignant lesions.

- **Eccentricity:** Indicates the elongation of the lesion. High eccentricity might be a sign of malignant growth, as cancerous lesions tend not to grow uniformly.

Area	Perimeter	Circularity	Bulkiness	Solidity	Eccentricity
11656	1836.788	0.043	0.290	0.290	0.578
11535	549.428	0.480	0.885	0.885	0.821
118975	5064.805	0.058	0.632	0.632	0.610
36321	570.191	1.404	0.729	0.729	0.851
128349	3837.975	0.109	0.804	0.804	0.170

Table II
Extracted Features (A)

B. Border Analysis [2]

Border Gradient Mean and Standard Deviation: These metrics assess the intensity changes at the edges of the lesion. Sharp gradients often suggest an abrupt transition between the lesion and normal skin, a typical feature of malignant lesions. Standard deviation provides a measure of how much the border's intensity varies, with higher variability often seen in cancerous lesions due to their uneven borders.

mean_Red	std_Red	mean_Green	std_Green	mean_Blue	std_Blue
1349.053	784.927	1356.667	792.586	1256.595	717.035
1380.908	1166.597	1362.470	1155.165	1353.406	1154.292
480.973	476.056	451.633	435.753	437.812	414.916
644.987	687.650	620.430	656.982	605.985	600.907
45.607	246.690	45.851	247.725	46.274	246.685

Table III
Extracted Features (B)

C. Color Features [4]

Mean and Standard Deviation of Each RGB Channel: The color of a lesion is a vital diagnostic indicator. Melanomas, for example, often display multiple shades of brown, black, red, and even blue and white, unlike benign moles which are typically uniform in color. The standard deviation reflects the diversity of color within the lesion, with higher values typically indicating potential malignancy.

mean_Red	std_Red	mean_Green	std_Green	mean_Blue	std_Blue
140.013	29.052	35.065	32.631	42.427	36.395
72.525	30.124	59.874	37.906	47.913	39.190
61.974	44.811	50.732	36.741	33.772	25.538
88.524	40.545	51.945	34.650	50.542	36.340
26.945	34.828	25.416	31.382	24.999	30.779

Table IV
Extracted Features (C)

D. Texture Features [1] [6]

- **Gray Co-occurrence Matrix (GLCM):** GLCM is a method for extracting second-order statistical texture features. It examines the frequency of co-occurring pixel values at a certain distance and orientation which helps

in identifying specific patterns of texture. Texture irregularities can be indicative of pathological changes within the tissue.
The extracted features where **contrast**, **dissimilarity**, **homogeneity**, **energy** and **correlation**.

contrast_0deg	contrast_45deg	contrast_90deg	contrast_135deg
248.262	312.322	163.537	302.570
106.850	143.977	105.367	153.593
187.893	232.956	151.797	249.264
58.000	85.305	53.738	77.847

Table V
EXTRACTED FEATURES (D)

- **Weber Local Descriptor:** This texture descriptor assesses local contrast, capturing subtle variations in intensity that might not be perceptible through basic imaging. High local contrast variability within a lesion can be a marker for malignancy as normal skin tends to have more uniform texture.

WLD_Mean	WLD_Std
0.03581	0.35469
0.00129	0.28212
0.00538	0.36684
0.00180	0.34379

Table VI
EXTRACTED FEATURES (D)

The extraction of these features plays a critical role in effectively differentiating between various types of skin lesions. Malignant lesions often exhibit greater asymmetry, border irregularity, color heterogeneity, and textural complexity compared to benign lesions. By quantitatively assessing these characteristics, machine learning models can be trained to detect subtle cues that predict the lesion's nature, enhancing diagnostic accuracy.

VI. MODEL SELECTION AND IMPLEMENTATION

This section details the evolution of the modeling strategies implemented to classify skin lesions as part of the Kaggle challenge. The approach was iterative, starting from a straightforward application of a deep learning model, moving through a hybrid model combining deep learning and traditional machine learning techniques, and culminating in an optimized ensemble model.

A. Initial Model Selection

The model architecture was designed to leverage both high-level visual features extracted by DenseNet201 and contextual information provided by metadata. This dual-input model ensures that decisions are informed by a comprehensive view of available data.

1) **Metadata Input:** A separate input layer was created to handle non-image data, consisting of various metadata features. This approach allows the model to use contextual information that may correlate with the classification targets.
2) **Feature Concatenation:** The features extracted from the DenseNet201 and those from the metadata input are combined into a single feature vector. This is achieved using a 'Concatenate' layer, ensuring that features from both sources can be jointly analyzed in subsequent layers.
3) **Additional Processing Layer:** After concatenation, the combined features are passed through a dense layer with 256 units and ReLU activation. This layer is designed to integrate and further process the combined information before making a final classification decision.
4) **Output Layer:** The final output layer uses a softmax activation function to classify the inputs into one of eight categories, corresponding to different types of skin lesions.

DenseNet201 is known for its efficiency in feature extraction due to its densely connected convolutional networks [3]. It is particularly effective for image classification tasks where preserving feature information throughout the network is crucial. Including metadata allows the model to consider additional factors that may not be discernible in the image data alone. This can include patient demographics, history, or other clinical parameters that are known to influence the diagnosis.

The decision to concatenate image and metadata features before further processing acknowledges the potential interplay between these data types. The additional dense layer after concatenation allows the model to learn interactions between these features, potentially leading to more accurate classifications. This architecture is justified by the need to handle complex data interactions in medical imaging tasks, where both visual cues and contextual information play critical roles.

*1) **Class Imbalance and Overfitting**:*

- *Class Imbalance:* Despite efforts to mitigate class imbalance through data augmentation, the initial imbalance might have influenced the model's training, potentially biasing it towards more frequently represented classes (class 2 had approximately 2000 more images than the other classes).

- *Overfitting:* The complexity of DenseNet201 makes it susceptible to overfitting, particularly when trained on a dataset not sufficiently large or diverse to support its deep architecture.

*2) **Hyperparameter Optimization**:* The lack of specific hyperparameter tuning for the dataset might have prevented the

model from achieving optimal performance. DenseNet201's default settings may not be the most effective for the classification task, highlighting the need for thorough optimization.

3) ***Feature Relevance****:* The autonomous feature extraction by DenseNet201 may not prioritize the most critical features needed to distinguish between different types of skin lesions effectively. This gap suggests a potential mismatch between the model's learning focus and the task-specific feature relevance.

4) ***Resource and Computational Constraints****:* Limited availability of computational resources significantly restricted training capabilities. The training was constrained to 30,000 image (approximately half of the available data) due to GPU memory limitations. This restriction likely impacted the model's ability to learn a comprehensive representation of the data, thus affecting its overall performance.

5) ***Evaluation Metrics****:* Relying solely on accuracy to evaluate the model's performance might not provide a complete picture, especially given the class imbalance. Alternative metrics that better account for the imbalance and the specific costs of misclassification may offer more insight into the model's effectiveness in a clinical setting.

B. Hybrid Model Implementation

To address the limitations observed with the initial DenseNet201 implementation, a hybrid modeling approach was developed. This approach involved integrating the predictions from the DenseNet201 model with those from a support vector machine (SVM), followed by a meta-classifier stage using logistic regression. This strategy aimed to leverage the strengths of both deep learning and traditional machine learning techniques to improve classification accuracy.

1) ***Stacked Model Configuration****:*

- *Integration of Predictions:* Predictions from the DenseNet201 and SVM models were combined to create a new feature set. This combination aimed to capture both the high-level features learned by the deep learning model and the nuanced decision boundaries defined by the SVM for the metadata.

- *Meta-classifier:* A logistic regression model was then trained on these combined predictions. The rationale behind using logistic regression as a meta-classifier was its effectiveness in weighting the predictions from different models, thereby refining the final decision process based on the strengths of each contributing model.

This model configuration not only provided a method to combine different types of predictions but also allowed for an additional layer of learning which refined the final output based on the validation set performance.

2) ***Improved Accuracy****:* The adoption of this hybrid model improved the classification accuracy to 52%, a notable increase from the initial results. This improvement underscores the potential of combining diverse modeling approaches to enhance performance, particularly in complex tasks such as medical image classification where different models may capture different aspects of the data.

3) ***Analysis of Hybrid Model Benefits****:*

- *Diverse Feature Representation:* By integrating features and predictions from different models, the hybrid approach likely benefited from a richer and more diverse representation of the data.

- *Error Correction:* The logistic regression stage provided a mechanism to correct individual model errors, fine-tuning the decision-making process based on the ensemble of predictions.

4) ***Challenges and Considerations****:*

- *Model Complexity:* While the hybrid model increased accuracy, it also introduced additional complexity in terms of model training and optimization.

- *Resource Requirements:* The need to train multiple models sequentially increased the computational overhead, requiring more careful management of resources.

The implementation of this hybrid model demonstrates the potential of ensemble techniques in tackling classification challenges where single models may fall short. Further refinement and optimization of this approach could yield even better results, making it a robust solution for skin lesion classification.

C. Final Model Optimization

Building on the hybrid model's success, the final optimization phase introduced several advanced techniques aimed at further enhancing the model's performance and addressing previous challenges. This phase focused on optimizing sample weighting, reducing feature dimensionality, and refining the neural network architecture.

1) ***Sample Weighting and DenseNet201 Adjustment****:*

- *Sample Weighting:* To improve learning efficiency and focus on difficult cases, the SVM's misclassifications were used to adjust the training samples' weights. This approach prioritized harder-to-classify examples, ensuring that these received more attention during the

training of the DenseNet201.

- *Model Refinement:* Some layers of DenseNet201 were unfrozen to allow fine-tuning on the specific features of the dataset. This was intended to make the model more responsive to the subtle nuances present in skin lesion images.

However, these adjustments introduced significant computational demands. The unfreezing of layers, while potentially beneficial for model accuracy, resulted in a substantial increase in the model's memory requirements. This led to difficulties in training due to the limited GPU resources, impacting the ability to train the model with the entire dataset.

*2) **Dimensionality Reduction with PCA**:* To manage the increased complexity and facilitate more efficient training, Principal Component Analysis (PCA) was implemented to reduce the dimensionality of the feature space.

PCA helped in compressing the features into a smaller set that retains 95% of the original variance, thus reducing computational load while preserving the essential information necessary for classification.

*3) **Enhanced Neural Network**:* The reduced feature set enabled the design of a more focused neural network, which was tailored to handle the dimensionally reduced data effectively. This neural network, featuring layers configured specifically for the task, was trained on the PCA-reduced features, enabling it to efficiently learn from the enhanced dataset. This process ensured that the model was not only fit on a representative subset of the data but was also validated against unseen data, providing a robust measure of its predictive capabilities.

*4) **Outcomes and Reflections**:* The final model optimization phase, while challenging due to resource limitations, resulted in an enhanced model that leveraged both deep learning adjustments and machine learning techniques to improve classification accuracy. These strategies, coupled with PCA for dimensionality reduction, offered a balanced approach to managing model complexity and computational feasibility.

D. Resource Constraints

Throughout the development of the skin lesion classification models, significant resource constraints were encountered that affected various stages of the project. These constraints primarily involved limited computational resources, such as GPU availability and memory, which posed challenges to model training and optimization efforts.

*1) **Impact on Initial Model Training**:*

- *Memory Limitations:* The initial deployment of the DenseNet201 model required substantial GPU memory due to the model's depth and complexity. The available hardware could not support training the model on the entire dataset, which consisted of approximately 62,000 images post-augmentation. As a result, the dataset was forced to be reduced to about 30,000 images, which likely compromised the model's ability to learn comprehensive and diverse features across all lesion types.
- *Training Time:* The complexity of DenseNet201 also extended training times significantly, limiting the ability to iterate quickly and experiment with different hyperparameter settings or architectural adjustments.

*2) **Effects on Hybrid and Final Model Implementations**:*

- *Complexity of Hybrid Model:* The hybrid model, which combined predictions from both DenseNet201 and an SVM, added further computational load. Training multiple models sequentially and managing the stacking process required careful allocation of resources to prevent memory overflows and ensure that each model component could be trained effectively.
- *Unfreezing DenseNet201 Layers:* In the final model optimization phase, unfreezing layers of DenseNet201 to fine-tune the model on specific features of the dataset further increased memory usage. This adjustment was intended to enhance the model's sensitivity to key diagnostic features but was hampered by GPU constraints, restricting the extent to which layers could be unfrozen and retrained.

*3) **Dimensionality Reduction as a Mitigative Strategy**:*

- *Implementation of PCA:* To counteract these issues, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the combined feature set from the DenseNet201 and SVM models. This reduction was crucial in decreasing the computational load, allowing for a more manageable training process on the available hardware.
- *Simplified Neural Network Architecture:* With the dimensionality reduced, a simpler neural network was designed that required less memory for training while still being effective in classifying the reduced feature set. This was crucial in making the best use of the limited resources.

VII. Results and Discussion

This section presents the outcomes of modeling efforts and discusses the implications of these results in the context of skin lesion classification. The performance of three model

iterations are compared: the initial DenseNet201 model, the hybrid model, and the final optimized model.

A. Model Performance Overview

Model	Accuracy	Precision	Recall	F1-Score
DenseNet201	67.1%	67.5%	66.9%	67.20%
Hybrid Model	79.98%	79.71%	79.90%	79.77%
Optimized Model	82.80%	84.37%	81.69%	83.0%

Table VII
PERFORMANCE METRICS FOR EACH MODEL ITERATION (MACRO-AVERAGED)

B. Discussion of Results

The initial DenseNet201 model showed promising capabilities but was hindered by its inability to effectively handle the diversity and complexity of the dataset under the constraints of limited computational resources. This was evident from its relatively lower performance metrics.

The hybrid model, integrating the DenseNet201 and SVM outputs, showed an improvement in accuracy, which reflects the benefit of combining diverse learning algorithms. The logistic regression meta-classifier effectively adjusted the weights of each model's predictions, enhancing overall decision accuracy.

The final model optimization, incorporating PCA and enhanced training strategies (such as focusing on misclassified examples), achieved the best performance. This underscores the effectiveness of targeted feature selection and the importance of focusing model training on challenging instances.

C. Model Implications and Practical Applications

The optimized model's improved accuracy and robustness indicate its potential for further improvement. Despite the challenges posed by resource limitations, the final model's use of PCA for dimensionality reduction and a simplified neural network structure demonstrates a viable pathway for deploying advanced machine learning models in environments with constrained computational resources.

D. Future Directions

Future work should focus on scaling the models to handle larger datasets without compromising performance, potentially through more efficient network architectures or advanced data sampling techniques. Exploring further ensemble techniques and advanced machine learning algorithms could yield better performance and robustness. Additional studies involving cross-validation across multiple datasets could help in verifying the model's generalizability and robustness in varied clinical scenarios.

VIII. CONCLUSION

This project lead to develop an effective classification system for skin lesions using advanced machine learning techniques. Through iterative modeling and optimization, the complexities of skin lesion classification were aimed to be addressed in a resource-constrained environment.

Starting with a basic application of the DenseNet201 model, which demonstrated limited success, model's accuracy and robustness was progressively improved through a series of strategic modifications. The integration of a hybrid model combining DenseNet201 and SVM outputs, followed by a logistic regression meta-classifier, enhanced classification accuracy. Subsequent optimizations, including sample weighting, PCA for dimensionality reduction, and neural network adjustments, further improved the model's performance, achieving the highest accuracy among the iterations.

A significant challenge faced throughout the project was the limitation of computational resources, which constrained the scope of data utilization and model complexity. Despite these limitations, strategic adjustments such as PCA and neural network simplification allowed us to effectively manage computational demands while still enhancing model performance.

The optimized model's performance suggests its potential utility in clinical settings for the early detection and classification of skin lesions. Its development within the constraints of limited resources underscores its applicability in various healthcare environments, not just those with access to high-performance computing facilities.

A. Future Research Directions

Future research should focus on enhancing the model's efficiency and scalability. Exploring lightweight neural network architectures or advanced training techniques could facilitate the deployment of similar models on a larger scale without requiring extensive computational resources.

Investigating other ensemble techniques and integrating more sophisticated machine learning algorithms could further refine classification accuracy and robustness.

Extending validation studies to include a broader range of clinical environments and demographic backgrounds would help ensure the model's generalizability and utility across diverse patient populations.

In conclusion, this project has demonstrated the potential of advanced machine learning techniques to significantly improve the classification of skin lesions, contributing to the fields of dermatology and oncology.

REFERENCES

[1] E. Claridge, P.N. Hall, and Others. The role of texture analysis in distinguishing pathological images. *Journal of Microscopy*, 168:203–212, 1992.

[2] H. Ganster, P. Pinz, and Others. Significance of border irregularity in dermatological image analysis. *IEEE Transactions on Medical Imaging*, 18(1):74–82, 1999.

[3] M. Naqvi, S. Q. Gilani, and Others. Comparison of deep learning models for classification of dermatological images. *Diagnostics*, 13(1911), 2023.

[4] D. A. Reddy, S. Roy, and Others. Analysis of geometric and color features in medical imaging. *Procedia Computer Science*, 22:1183–1192, 2013.

[5] World Health Organization. Radiation: Ultraviolet (uv) radiation and skin cancer, 2024.

[6] M. M. Yamunadevi, S. Siva Ranjani, and Others. Efficient segmentation of the lung carcinoma by adaptive fuzzy-glcm with deep learning based classification. *ResearchGate*, 2020.