Dermoscopic Image Classification IMA205 - Challenge

Daniel Felipe TORRES ROBLES
daniel.torresrobles@telecom-paris.fr

Télécom Paris

***Abstract*—This report details our approach to the Skin Lesion Classification Challenge. The challenge involves classifying dermoscopic images of skin lesions into eight diagnostic classes. Our methodology includes lesion segmentation using a revisited OTSU algorithm, feature extraction with a ResNet152 model, and classification with an SVM. Results demonstrate the effectiveness of this approach, achieving competitive performance on the test dataset.**

***Index Terms*—Skin Lesion, Classification, ResNet152, SVM, OTSU Thresholding**

I. Introduction

A skin lesion is defined as a superficial growth or patch of the skin that is visually different and/or has a different texture than its surrounding area. Skin lesions, such as moles or birthmarks, can degenerate and become cancer, with melanoma being the deadliest skin cancer.

Malignant melanoma is one of the leading cancers among many white-skinned populations globally. Changes in recreational behavior, together with increased ultraviolet radiation exposure, have dramatically increased the number of melanomas diagnosed in recent decades. The incidence of melanoma was first noted in the United States in 1930, where one person out of 100,000 suffered from skin cancer. This rate increased to six per 100,000 in the 1950s and to 13 per 100,000 in 1991. The figures are comparable to those observed in Europe. In 1995, Austria had an incidence rate of 12 per 100,000, reflecting a 51.8% increase over the previous ten years. The incidence of melanoma continues to rise, highlighting the urgent need for effective detection and treatment strategies [5].

Fortunately, investigations have shown that early detection can significantly improve survival rates, as melanoma is nearly 100% curable if recognized early and treated surgically. While the mortality rate in the 1960s was around 70%, today's survival rate is approximately 70%, primarily due to early detection [5].

A. Automated Melanoma Diagnosis

The high incidence of malignant melanoma has led to growing interest in automated diagnosis. Several studies have focused on automated melanoma recognition using image processing techniques. Unfortunately, complete integrated dermatological image analysis systems are rare. However, promising developments in computer-aided diagnosis (CAD) systems offer a glimpse into future solutions [5].

B. Goal of the Challenge

This challenge aims to classify dermoscopic images of skin lesions into eight different diagnostic classes:

1) Melanoma
2) Melanocytic nevus
3) Basal cell carcinoma
4) Actinic keratosis
5) Benign keratosis
6) Dermatofibroma
7) Vascular lesion
8) Squamous cell carcinoma

The dataset provided for the challenge contains 25,331 dermoscopic images, with segmentation maps and metadata available for some images. The data is divided into training-validation (75%) and test (25%) sets. Our goal is to estimate the correct class of each dermoscopic image in the test set, using only the provided data [1].

C. Related Work

Several studies have tackled the automated classification of skin lesions, providing valuable insights into the basis for our proposed work:

- *Conditional Dependence Tests:* Reimers et al. revealed that deep learning models could identify features associated with the ABCD rule and bias variables without explicit feature extraction. This study is foundational in showing that deep learning models inherently capture ABCD rule patterns, even without explicit feature engineering [2].

- *ResNet-SVM Fusion:* Sahli et al. improved classification performance by combining features from the ResNet model with a Support Vector Machine (SVM) classifier for glioblastoma tumor segmentation and classification. Their innovative approach of removing the final ResNet layer and leveraging the SVM classifier demonstrated the power of feature fusion, significantly enhancing classification accuracy. This concept is directly applied in our work to classify skin lesions [3].

- *Otsu Thresholding for Segmentation:* Tom and Daba revisited the Otsu algorithm for skin cancer segmentation, achieving high accuracy in lesion segmentation. Their modified Otsu method ensures precise lesion boundary identification, laying a strong foundation for accurate feature extraction. This algorithm is critical to our lesion segmentation approach, providing reliable segmentation maps for classification [4].

These three studies form the basis of our proposed work, highlighting the potential of computer-aided diagnosis in dermatology. Additionally, insights from other works were considered, such as automated melanoma recognition techniques [5] and the comprehensive analysis of pigmented skin lesions [6]. Work on deep neural networks for dermatologist-level classification of skin cancer [9], and a computational approach for detecting pigmented skin lesions [7] were also influential in shaping our methodology.

II. Data Preprocessing

The dataset consists of a total of 25,331 dermoscopic images of skin lesions, used in separate training and test sets:

- Number of images in the training set: 18,998
- Number of images in the test set: 6,333
- Number of segmentation masks in the training set: 1,945
- Number of segmentation masks in the test set: 648

A. Missing Metadata

Some metadata, such as age, sex, and anatomical position, are available for the majority of images. However, there are missing data points:

- In the training metadata:
  - Missing data in 'SEX': 284 (1.49%)
  - Missing data in 'AGE': 324 (1.70%)
  - Missing data in 'POSITION': 1,970 (10.37%)
- In the test metadata:
  - Missing data in 'SEX': 100 (1.58%)
  - Missing data in 'AGE': 113 (1.78%)
  - Missing data in 'POSITION': 661 (10.44%)

Despite these missing values, the data available is sufficient for meaningful analysis and modeling.

B. Segmentation Masks

The provided folders contain a mix of images and segmentation masks. However, not all images have corresponding masks. To facilitate processing, we separated the images and masks into different folders, ensuring clear identification.

C. Data Imbalance

As seen in Figure 1, certain classes are vastly underrepresented compared to others. For instance, the Melanocytic nevus class (Class 2) dominates, while classes like Dermatofibroma and Vascular lesion have very few samples. This imbalance presents a challenge as classifiers can become biased towards the dominant classes, leading to suboptimal performance on underrepresented ones.

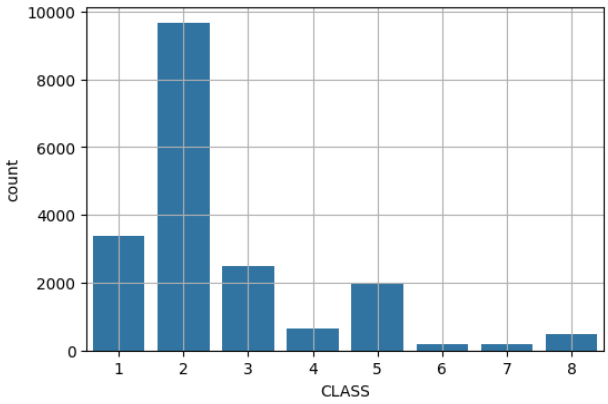

Fig. 1. Distribution of Classes in Training Data

To address this imbalance issue, we could employ the following strategies:

- *Data Augmentation:* Increase the number of samples in underrepresented classes through transformations like rotation, flipping, and scaling.
- *Class Weighting:* Assign higher misclassification penalties to minority classes to balance the training.

These strategies aim to ensure fairer representation and improved classifier performance across all classes.

D. Image Cropping

The dataset includes some squared images taken through a circular lens, leading to the primary region of interest being confined within a central circle. To effectively focus our model on these areas, a specialized mask is employed to isolate this circular region. This mask is a binary image designed to emulate the lens pattern, with black regions in the rounded corners and a white central circle.

1) Motivation: Using this mask, we can establish correspondence with all images in the dataset and subsequently crop them based on a specified coefficient to ensure the model is trained primarily on the regions of interest. This technique enhances the precision of our image processing pipeline.

2) Implementation: The image cropping process involves two key steps:

1. Calculating the Coefficient of Correspondence:
The coefficient of correspondence measures how well the circular region aligns with the circular lens mask. This coefficient is calculated by comparing each image against the circular mask template and normalizing the results to a value between 0 and 1.

2. Cropping Images Based on the Coefficient:
Each image is cropped to focus on the region of interest using the calculated coefficient. The cropping size varies depending on the coefficient value, ensuring optimal attention to the circular region. The cropping is defined as follows:

- Coefficient ≥ 0.95: Larger cropping area
- Coefficient < 0.95 and ≥ 0.6: Medium cropping area
- Coefficient ≤ 0.6: No cropping

3) Batch Processing: The images are processed in batches, iterating through a DataFrame containing image metadata. Each image is checked for cropping applicability, and the appropriate cropping is applied based on the coefficient value.

4) Example Cropped Image: The following image illustrates a cropped dermoscopic image focusing on the primary region of interest.

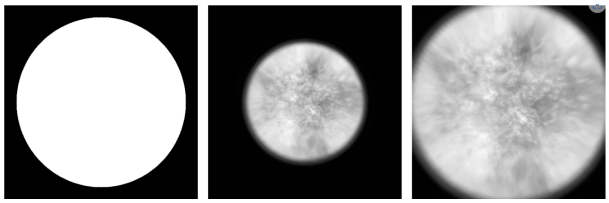

Fig. 2. Example of a Cropped Dermoscopic Image

E. OTSU Thresholding

1) Motivation: OTSU thresholding is an effective image processing technique for automatically converting grayscale images into binary form by distinguishing the foreground from the background. It accomplishes this by analyzing the histogram of the image to find a threshold that minimizes the variance within each class while maximizing the variance between classes. This method is particularly advantageous when the image features two distinct pixel classes, such as in medical imaging where it can differentiate between affected and healthy tissue.

Following the cropping process described earlier, where background elements outside the lens are removed, OTSU thresholding becomes even more valuable. It helps to further clarify and enhance the distinction between significant features, such as skin and lesions, rather than simply differentiating between the lens boundaries and the interior. This refined focus aids in improving the accuracy of subsequent image analysis and model training tasks.
[4]

2) Implementation: The image segmentation process using OTSU thresholding involves three main steps:

1. Preprocessing: Images are initially converted to grayscale and Gaussian smoothing is applied to reduce noise and improve threshold accuracy.

2. Applying OTSU's Algorithm: The revisited OTSU algorithm is used to convert the grayscale images into binary form, separating the foreground (lesion) from the background (skin). This is achieved by analyzing the histogram to identify the optimal threshold value.

3. Mask Application: Once the images have been segmented, the resulting binary masks are inverted to focus on the lesion regions. These masks are then applied to the original images to isolate the lesions effectively.

Batch Processing of Images: After creating a directory structure for organized storage, all images in the dataset are processed using the OTSU thresholding technique. Cropped images are read from the source directory, processed, and the binary masks are saved to a separate directory. The segmented images are then generated by applying the binary masks to the original images.

Segmentation Results: The segmentation results using the revisited OTSU algorithm are illustrated in Figure **??**. On the left is an original dermoscopic image containing a skin lesion, while on the right is the corresponding segmented image. The OTSU thresholding technique effectively isolates the lesion, distinguishing it clearly from the surrounding healthy skin.

The binary mask generated from the OTSU algorithm accurately identifies the lesion boundary, enabling the creation of a clean and focused segmentation map. This refined segmentation enhances the clarity of features for subsequent classification tasks, ensuring that our deep learning model primarily focuses on the region of interest.

Fig. 3. Segmented Dermoscopic Images using OTSU Thresholding

F. Data Augmentation

Data augmentation is an essential technique that enhances the diversity of the training dataset by applying various transformations, such as resizing, cropping, and normalization. Leveraging the 'torchvision' library, this process generates new variations of images, thereby improving the model's generalization ability and performance by exposing it to a broader range of visual conditions.

1) Motivation: The dataset is significantly imbalanced across the eight diagnostic classes, leading to underrepresentation of certain classes. Data augmentation helps mitigate this issue by artificially increasing the number of samples in the minority classes. By balancing the dataset, we aim to improve the model's ability to recognize all classes effectively.

2) Implementation: To implement data augmentation, we followed these steps:

1. Dataset Preparation: - Metadata Loading: Metadata was loaded from a CSV file containing image IDs and class labels. - *Organizing Directories:* Separate directories were created for each class in both training and validation datasets. - *Dataset Splitting:* The dataset was split into training and validation sets, ensuring stratification to maintain class balance.

2. Augmentation Transformations: Several augmentation transformations were applied to enhance the dataset: - *Horizontal Flip:* Randomly flips the image horizontally. - *Rotation:* Rotates the image by a random angle within a specified range. - *Color Jitter:* Adjusts brightness, contrast, saturation, and hue randomly. - *Resized Crop:* Crops the image to a random size and aspect ratio, then resizes it to a fixed dimension. - *To Tensor:*Converts the image to a tensor for processing by neural networks.

3. Custom Dataset Class: A custom dataset class, 'AugmentDataset', was developed to apply the augmentation transformations based on predefined factors. Each class was assigned a specific factor, representing the number of augmented samples required.

4. Saving Augmented Images: A function was created to save the augmented images into their corresponding directories, ensuring that the original images were also retained.

Augmentation Process Overview: 1. *Horizontal Flip:*Improves recognition of skin lesions by generating horizontally flipped versions of images. 2. *Rotation (30 degrees):*Simulates lesions at different orientations. 3. *Color Jitter:*Introduces variations in brightness, contrast, and hue, mimicking different lighting conditions. 4. *Resized Crop:*Provides diverse perspectives by cropping and resizing images to fixed dimensions.

3) Results: The augmentation factors and their corresponding number of images are summarized in Table I. The same augmentation process was applied to the validation set.

TABLE I
DATA AUGMENTATION RESULTS PER CLASS

Class	Original Images	A. Factor	Augmented Images
1	2713	2	5426
2	7725	1	7725
3	1994	3	5982
4	520	14	7280
5	1574	4	6296
6	143	54	7722
7	152	50	7600
8	377	20	7540

By applying these data augmentation techniques, we significantly enhanced the diversity and balance of the training dataset, improving the model's ability to generalize across different classes and visual conditions.

III. RESNET-152 MODEL

A. Overview and Key Features

ResNet-152 is a state-of-the-art deep learning model that leverages a deep architecture of 152 layers to handle the complexity of dermoscopic images. It incorporates residual blocks, which facilitate effective training by preventing vanishing gradients, a common issue in very deep networks. This structure allows ResNet-152 to accurately extract intricate features related to the Asymmetry, Border irregularity, Color, and Dimension (ABCD) of lesions, crucial for melanoma detection.
[2]

1. *Deep Learning Capabilities:* - The model's architecture is exceptionally well-suited for handling complex dermatoscopic images due to its ability to identify subtle variations in skin lesions that could indicate different skin conditions or cancers.

2. *Residual Blocks for Effective Training:* - Residual blocks facilitate effective training of deep networks by allowing gradients to flow smoothly through the network layers, preventing degradation in performance even as the network depth increases.

3. *Handling Diverse Data:* - ResNet-152 can generalize from complex patterns, making it adept at differentiating between various categories of skin lesions effectively, even when trained on a dataset with varied and nuanced features.

4. *Pre-training and Fine-tuning:* - Pre-training on large-scale datasets like ImageNet and subsequent fine-tuning on dermatoscopic images significantly enhances the learning process.

B. Training Analysis

The training and validation process of ResNet-152 highlighted several important aspects of its performance in classifying skin lesions:

1. *Loss and Accuracy Trends:* - During training, both the training and validation loss showed a consistent downward trend, indicating effective learning. Validation accuracy reached over 70%, demonstrating the model's capability to generalize well.

2. *Early Stopping and Generalization:* - Early stopping based on validation loss prevented overfitting and ensured the model remained robust across different skin lesion classes. The model maintained high accuracy across the majority of classes.

3. *Residual Blocks Contribution:* - The residual blocks played a crucial role in stabilizing the training of the 152-layer network, preventing gradient vanishing and ensuring efficient gradient flow.

C. Prediction on the Test Set

ResNet-152's ability to predict on the test dataset was further explored through probabilistic analysis. By removing the final classification layer, the model provided rich feature representations and probabilities for each lesion, enabling detailed probabilistic insights.

Key Observations:

1. *Probabilistic Distribution:* - The probabilistic distribution across classes reflected the nuanced differences between skin lesion categories, with a high confidence score for the correct class in most cases.

2. *Misclassification Patterns:* - Analysis of misclassified lesions revealed that lesions with similar features to others (e.g., benign keratosis vs. actinic keratosis) were often confused, highlighting the importance of further refinement in feature extraction.

3. *Class Imbalance Impact:* - Despite data augmentation, class imbalance still had an impact, particularly for underrepresented classes like vascular lesions. This necessitates additional strategies such as class weighting during training.

D. Future Work

The ResNet-152 model, with its residual blocks and deep architecture, is well-suited for classifying skin lesions. By leveraging pre-training and fine-tuning, it achieved high accuracy across various classes. However, further improvements can be made by addressing the impact of class imbalance and enhancing the feature extraction process for nuanced lesions. Future work will focus on refining the probabilistic analysis and exploring hybrid models that combine deep learning with traditional machine learning approaches.

E. Confusion Matrix Analysis

A confusion matrix provides insight into the performance of a classification model by comparing the predicted and true classes. In the context of this challenge, it reveals how well the ResNet model distinguishes between the eight diagnostic classes of skin lesions. Figure 4 shows the confusion matrix obtained from the predictions on the training set.

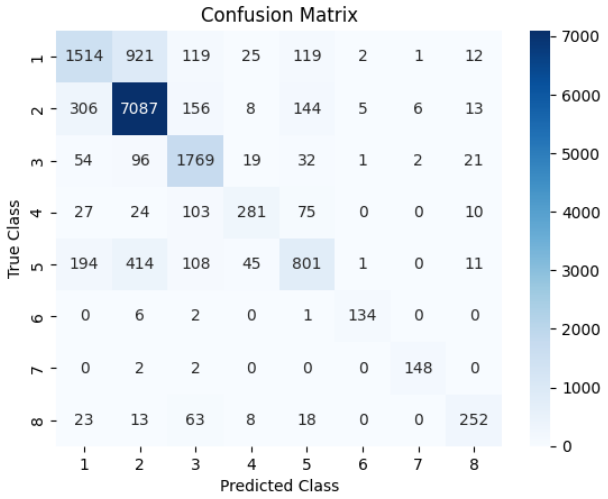

Fig. 4. Confusion Matrix Comparing True and Predicted Classes using ResNet Model

1) Analysis of the Confusion Matrix: From the confusion matrix, several key observations can be made:

1. *Dominance of Melanocytic Nevus (Class 2):* - The Melanocytic Nevus class has the highest number of correctly classified images (7,087), which is expected due to its significant representation in the training data. - However, it is also often confused with Melanoma (Class 1), leading to a relatively high number of false positives (306 instances).

2. *Melanoma (Class 1):* - The model accurately classified Melanoma in 1,514 cases but confused it with Melanocytic Nevus in 921 instances. - This highlights the challenge of distinguishing between these two classes due to their visual similarities.

3. *Basal Cell Carcinoma (Class 3):* - Basal Cell Carcinoma is accurately classified in 1,769 cases. - It is occasionally confused with Actinic Keratosis (Class 4), which might be attributed to overlapping visual features.

4. *Actinic Keratosis (Class 4):* - Actinic Keratosis has moderate classification accuracy, with 281 correctly predicted instances. - However, it is confused with Benign Keratosis (Class 5) and Basal Cell Carcinoma (Class 3), indicating some overlap in features.

5. *Benign Keratosis (Class 5):* - Benign Keratosis is classified correctly in 801 cases, but the model struggles to differentiate it from Actinic Keratosis (Class 4), leading to 108 misclassifications.

6. *Dermatofibroma (Class 6):* - Dermatofibroma is classified with high accuracy due to its distinct visual characteristics. - There are very few misclassifications, as it is visually distinguishable from other lesions.

7. *Vascular Lesion (Class 7):* - Vascular Lesion is also classified accurately, with minimal confusion with other classes.

8. *Squamous Cell Carcinoma (Class 8):* - Squamous Cell Carcinoma is classified with moderate accuracy (252 correct predictions), but the model confuses it with Basal Cell Carcinoma (Class 3) and Melanoma (Class 1).

The confusion matrix analysis highlights that while the ResNet model effectively distinguishes between most skin lesion classes, there are challenges in differentiating between visually similar classes, such as Melanoma vs. Melanocytic Nevus and Benign Keratosis vs. Actinic Keratosis. This difficulty underscores the importance of incorporating additional data features like metadata (age, sex, and anatomical position) to correct biases and improve classification accuracy.

IV. Support Vector Machine (SVM)

A. Motivation

The ResNet model provides detailed feature representations and class probabilities that can be leveraged for improving classification accuracy. However, certain visually similar classes, such as Melanoma vs. Melanocytic Nevus and Benign Keratosis vs. Actinic Keratosis, are often misclassified due to overlapping features. Incorporating metadata (age, sex, anatomical position) can help correct biases and refine the classification. An SVM model, using the ResNet features and metadata, aims to address these challenges by better differentiating between these visually similar classes.
[3]

B. Preprocessing and Feature Engineering

The feature set used to train the SVM model includes probabilities obtained from the ResNet model, along with metadata features. The preprocessing steps involved:

1. *Data Cleaning:* - Removal of irrelevant columns like ‘ID‘ and ‘Predicted Class‘. - Filling missing values with zero.

2. *Categorical Encoding:* - Conversion of categorical variables (‘SEX‘ and ‘POSITION‘) into numerical features using one-hot encoding.

3. *Feature Scaling:* - Scaling of numerical features (‘AGE‘ and class probabilities) using standard normalization techniques.

These preprocessing steps ensure that the SVM model receives a consistent and normalized dataset, which aids in better classification performance.

C. Model Training and Optimization

A Support Vector Machine (SVM) model was chosen for classification due to its ability to handle high-dimensional data. To identify the best hyperparameters, a grid search was conducted. This method systematically explores a predefined hyperparameter space, evaluating model performance using cross-validation to find the best combination of parameters.

Hyperparameter Tuning: - *Kernel:*Polynomial kernel (‘’poly‘‘). - *Class Weight:*Balanced to address class imbalance. - *Gamma:*Explores different values to control the kernel’s influence. - *C:*Determines the trade-off between classification accuracy and margin width.

Optimal Parameters Found: After the grid search, the best parameters identified for the SVM model were: - *Kernel:*‘’poly‘‘ - *Gamma:*‘0.5‘ - *C:*‘10‘

D. Evaluation of the SVM’s Impact

To evaluate the impact of the SVM model, a confusion matrix was plotted for the validation dataset predictions. This matrix provides insights into how well the SVM model corrects the misclassifications from the ResNet model.

Confusion Matrix Analysis: 1. *Reduced Misclassifications:* - The SVM model significantly reduces misclassifications for visually similar classes like Melanoma vs. Melanocytic Nevus. - For example, Melanoma (Class 1) misclassifications with Melanocytic Nevus (Class 2) decreased substantially.

2. *Metadata Integration Benefits:* - Incorporating metadata such as age, sex, and anatomical position allows the SVM model to differentiate between lesion types more effectively, correcting biases observed in the ResNet model. - For instance, Benign Keratosis vs. Actinic Keratosis (Classes 5 and 4) misclassifications were reduced due to metadata use.

3. *Improved Classification Accuracy:* - The overall classification accuracy improved due to the refined feature set and metadata integration. - Incorporating metadata helped the

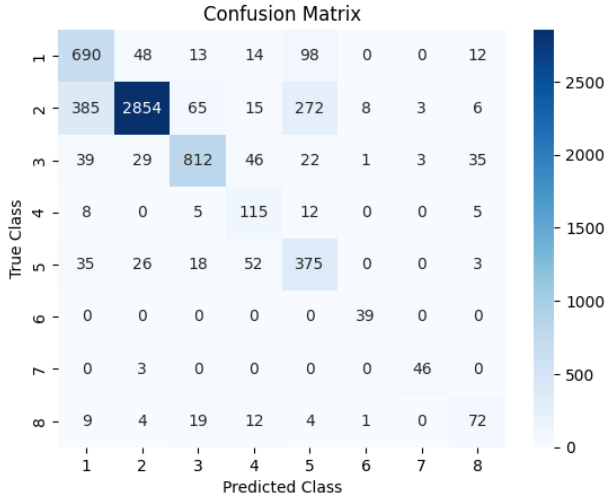

Fig. 5. Confusion Matrix of SVM Model on Validation Dataset

SVM model achieve better discrimination between nuanced lesion features.

By leveraging probabilities obtained from the ResNet model, along with metadata, the SVM model corrects misclassifications and improves overall classification performance. Grid search optimization ensured that the SVM model was configured for optimal predictive performance across all diagnostic classes. This approach demonstrates the importance of combining deep learning feature representations with traditional machine learning classifiers to enhance skin lesion classification accuracy.

V. Final Evaluation and Improvements

A. Kaggle Test Set Evaluation

The final classification performance of the ResNet-SVM model was evaluated using the test images on the Kaggle platform. The model achieved a score of 61%, demonstrating reasonable accuracy but leaving significant room for improvement. This result is indicative of the complexity inherent in differentiating between the eight diagnostic classes of skin lesions and the challenges posed by the high visual similarity of some classes.

1) Challenges and Observations: 1. *Visual Similarity and Overlapping Features:* - Some skin lesions, such as Melanoma and Melanocytic Nevus, share overlapping visual features, making them challenging to classify correctly. - Similarly, benign and malignant keratoses are often misclassified due to their similar appearance.

2. *Metadata Limitations:* - Missing metadata, particularly in anatomical position and patient sex, limited the ability of the model to fully leverage this information for classification correction. - Incorporating more comprehensive metadata could enhance differentiation between visually similar lesions.

3. *Class Imbalance Impact:* - Despite data augmentation efforts, class imbalance still impacted classification performance, particularly for underrepresented classes like vascular lesions.

2) Areas for Improvement: 1. *Hyperparameter Optimization:* - Further optimization of the hyperparameters in the ResNet and SVM models could yield significant improvements. While grid search was utilized for the SVM model, a more exhaustive search might reveal better configurations.

2. *Feature Engineering Refinement:* - More detailed preprocessing and feature engineering could improve the quality of input features for the SVM model. For example, generating additional metadata-based features could provide better context for classification.

3. *Training Data Expansion:* - Expanding the training dataset by incorporating external annotated datasets could provide additional lesion variations, helping the model generalize better.

4. *Hybrid Models and Ensemble Learning:* - Combining multiple classifiers or leveraging ensemble learning techniques might help mitigate biases inherent in any single model, improving overall performance.

The final Kaggle score of 61% indicates that the ResNet-SVM model provides a solid baseline for automatic skin lesion classification. However, further improvements can be made by focusing on parameter adjustment in each step of the process.

References

[1] IMA205 Challenge 2024. Available at https://www.kaggle.com/competitions/ima205-challenge-2024

[2] Reimers, M., Lang, T., Biester, M., de la Torre, P., Gessert, N., Schlaefer, A., and Baumgartner, D. (2021). Conditional Dependence Tests Reveal the Usage of ABCD Rule Features and Bias Variables in Automatic Skin Lesion Classification. Available at https://typeset.io/papers/conditional-dependence-tests-reveal-the-usage-of-abcd-rule-5eihvfh90j

[3] Sahli, R., Tamer, M., and Hamdi, R. (2022). ResNet-SVM: Fusion Based Glioblastoma Tumor Segmentation and Classification. Available at https://typeset.io/papers/resnet-svm-fusion-based-glioblastoma-tumor-segmentation-and-aa6r1hvl

[4] Tom, F., and Daba, M. (2023). Revisited Otsu Algorithm for Skin Cancer Segmentation. Available at https://typeset.io/papers/revisited-otsu-algorithm-for-skin-cancer-segmentation-3fto4nhe

[5] Celebi, M. E., Iyatomi, H., Schaefer, G., and Stoecker, W. V. (2009). Automated melanoma recognition. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4–7. http://ieeexplore.ieee.org/document/918473

[6] Korotkov, K., and Garcia, R. (2012). Computerized analysis of pigmented skin lesions: A review. *Computers in Biology and Medicine*, 56, 92–114. https://www.sciencedirect.com/science/article/pii/S0933365712001108#bib0180

[7] González-Díaz, I. (2017). A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications*, 85, 331–341. https://www.sciencedirect.com/science/article/pii/S0957417416302354#bib0023

[8] Menzies, S. W., Kreusch, J., Byth, K., Pizzichetta, M. A., Braun, R. P., Malvehy, J., and Zalaudek, I. (2014). Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. *Journal of the American Academy of Dermatology*, 70(4), 613–621. https://www.sciencedirect.com/science/article/pii/S0933365713001589

[9] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118. https://www.nature.com/articles/nature21056

[10] ElMoufidi, I., and Mahmoudi, S. (2016). Classification of Melanoma Lesions Using Sparse Coded Features and Random Forests. Available at https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01250955/document

[11] Sixth ISIC Skin Image Analysis Workshop @ CVPR 2021 Virtual. Available at https://workshop2021.isic-archive.com/

[12] Research Datasets for Skin Image Analysis. Available at https://workshop2020.isic-archive.com/#paper