Dermoscopic Image Classification Challenge: Leveraging Machine Learning for Early Detection of Skin Cancer

Adnane El Bouhali
Télécom Paris
IMA205

May 2024

Abstract

This report presents the methodologies and findings from the dermoscopic image classification challenge aimed at identifying skin cancer types from photographic images. Skin lesions vary by visual and textural characteristics, which can be precursors to malignancies such as melanoma—the deadliest form of skin cancer. Through non-invasive computer-aided diagnosis (CAD), early detection can significantly improve treatment outcomes. This challenge focuses on the classification of dermoscopic images into eight distinct categories using machine learning techniques, highlighting the importance of accurate early detection in dermatological practice.

0.1 Introduction

In the rapidly evolving fields of dermatology and oncology, the integration of technology has provided transformative solutions for early cancer detection and treatment.

Background: A skin lesion is defined as a superficial growth or patch of the skin that differs visually and/or texturally from its surrounding area. Such lesions, including moles and birthmarks, can degenerate into cancerous forms, with melanoma recognized as the most lethal type of skin cancer. Over recent decades, the incidence of melanoma has escalated, particularly in populations with significant exposure to ultraviolet radiation.

Importance of Early Detection: Early detection and surgical excision remain the most effective treatments for skin cancer. The advancement of non-invasive computer-aided diagnosis (CAD) systems has marked a pivotal shift in dermatological practice, significantly enhancing the early identification and management of skin lesions suspected of malignancy.

Challenge Objective: The primary goal of this challenge is to classify dermoscopic images into one of eight diagnostic categories using machine learning algorithms. These categories are:

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

To achieve this, participants will employ the ABCD rule—evaluating Asymmetry, Border irregularity, Color, and Dimension of each lesion—to extract relevant features from the images. These features, combined with sophisticated machine learning techniques, will form the basis for image classification.

0.2 Dataset Overview

0.2.1 Description of the Dataset

The dataset provided for this challenge consists of 25,331 dermoscopic images of skin lesions, accompanied by relevant segmentation data and metadata when available. The images are pre-divided into a training-validation set (75%) and a test set (25%), with only the training-validation set's classifications made by clinicians available to the participants.

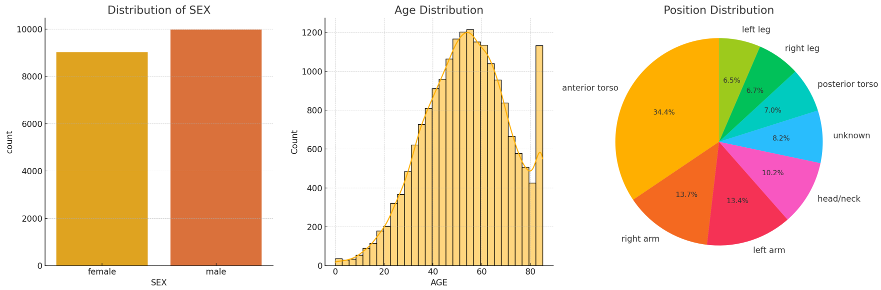

Figure 1: Exploratory Data Analysis (EDA) plot.

0.2.2 Challenges Presented by the Dataset

The main challenges posed by the dataset include:

- **Limited Segmentation Data:** Only about 10% of the images have accompanying ground-truth segmentations, which are crucial for accurate feature extraction using traditional methods.

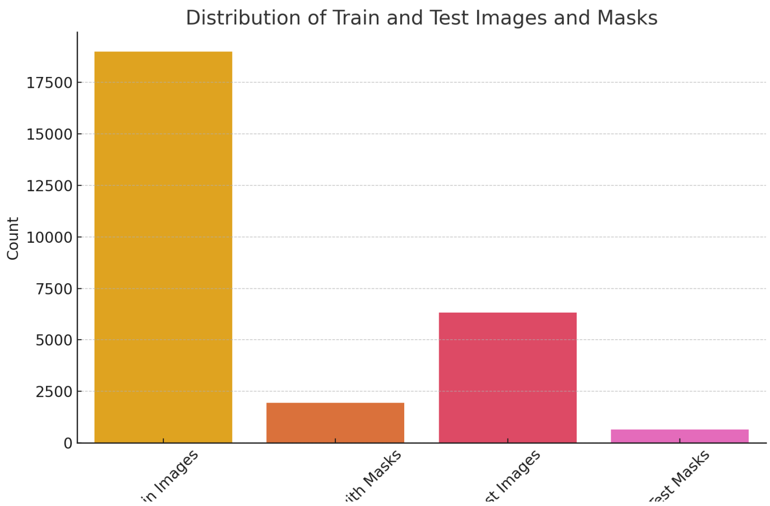

Figure 2: Distribution of Images With Masks.

- **Missing Metadata:** Approximately 2,000 images lack complete metadata, including patient age, sex, and anatomical position, which are important for enhancing classification accuracy.

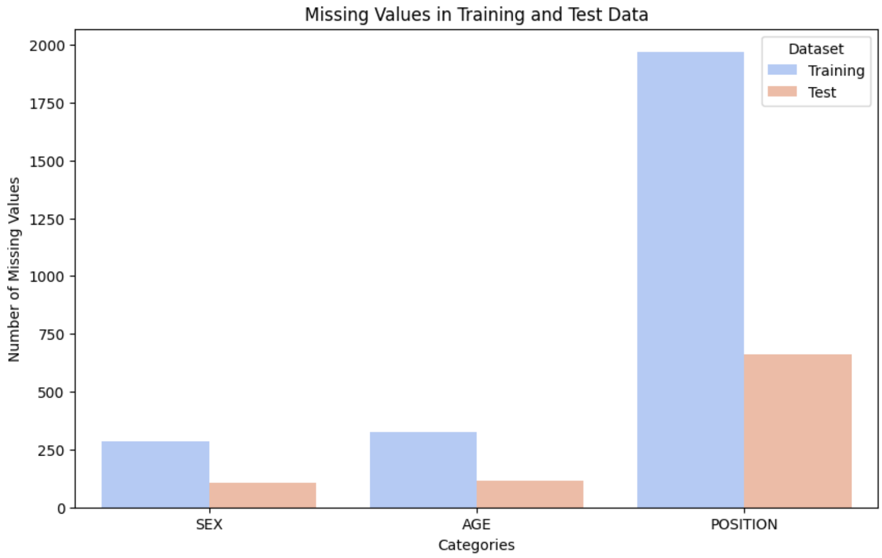

Figure 3: Missing Values in the Dataset.

- **Class Imbalance:** The dataset exhibits significant imbalances among the different diagnostic categories, which can bias the performance of machine learning models towards the more frequently represented classes.

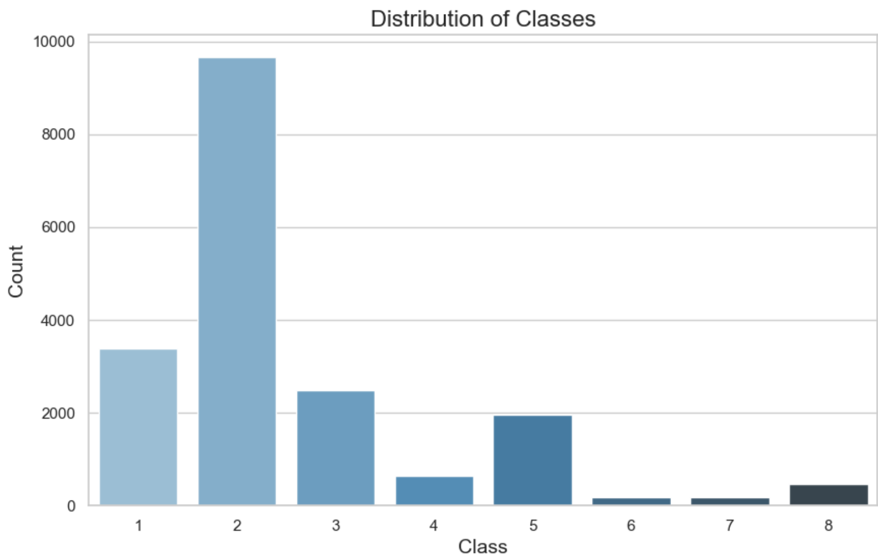

Figure 4: Class Distribution

0.2.3 Proposed Solutions to Address the Challenges

To effectively manage these challenges, the following strategies were adopted:

- **Data Augmentation:** To mitigate the challenges posed by the inherent imbalance within the dataset, several data augmentation strategies were implemented. One of the key techniques utilized was the Adaptive Synthetic Sampling Approach (ADASYN). This method is particularly effective in generating synthetic samples for minority classes, thereby equalizing the distribution across different classes and enhancing the robustness of the dataset. Figure 5 vividly depicts the class distribution of the training dataset post the application of these preprocessing and augmentation measures. This visualization highlights the improved balance among the classes, showcasing the effectiveness of the data augmentation in addressing the initial disparities.

Figure 5: Improved Balance Among the Classes After Data Augmentation.

- **Imputation of Missing Data:** In addressing the issue of missing metadata within our dataset, we evaluated multiple strategies to handle the absence of data points effectively. One of the initial strategies considered was the complete elimination of records that contained missing data. However, this approach was quickly ruled out due to the significant volume of data that would be lost, which was deemed too substantial relative to the overall dataset size.

  As an alternative, for missing numerical data such as age, we opted to use the median age derived from the training dataset as a substitute value. This method helps maintain statistical integrity without skewing the data distribution significantly.

For missing categorical data, such as class and position labels, we introduced a systematic approach by assigning a placeholder value labeled "unknown." This step was essential to preserve the structure of the dataset while clearly marking data as unprovided. To facilitate analysis and maintain data integrity, these placeholder values were processed through one-hot encoding. However, to prevent the introduction of unnecessary complexity and potential model bias, columns that corresponded solely to "unknown" values were subsequently removed from the dataset.

This strategy was designed to gracefully handle missing entries by acknowledging their absence, thus avoiding the pitfalls of introducing artificial noise or distortions through arbitrary data substitutions. The goal was to preserve the original dataset's fidelity while providing a robust foundation for further analysis and modeling.

These strategies aim to maximize the efficacy of the machine learning models by enhancing the quality and balance of the dataset, thus ensuring fair and robust classification across all categories.

0.3 Segmentation Strategy

For handling the challenge of missing segmentations within our dataset, we turned to a specialized deep learning model. The model employed, a modified version of the ResNet architecture, was developed specifically for segmentation tasks, known as MFSNet. This innovative architecture was meticulously trained using all available ground-truth segmentations, which served as a solid foundation for the model's learning process.

Once adequately trained, MFSNet was tasked with generating segmentations for those images lacking this essential data. This was complemented by an additional preprocessing step involving an in-painting script, which proved instrumental in minimizing the impact of occluding elements like hair on the quality of the segmentations.

The choice of MFSNet, despite its higher computational demands and slower operation compared to more straightforward methods such as OTSU thresholding, was driven by its superior ability to produce high-quality segmentations. These high-quality outputs are crucial for subsequent accurate feature extraction and classification, highlighting the importance of effective segmentation in enhancing overall model performance. This approach not only compensates for the missing data but also fortifies the dataset, enabling more reliable and detailed analyses.

0.4 Feature Extraction

The process of feature extraction was meticulously designed to harness both the intrinsic and the perceptible characteristics of the skin lesions. Utilizing the ABCD rule as our foundational guide—referring to Asymmetry, Border, Color, and Dimension—we manually extracted features based on these criteria. The features were strategically selected based on the insights provided by Ganster et al, who have laid out a comprehensive framework for analyzing these aspects in dermatoscopic images.

To further refine our feature set, we employed advanced textural analysis techniques. The Gray Level Co-occurrence Matrix (GLCM) and the Weber Local Descriptor (WLD) were instrumental in capturing the textural nuances of the skin lesions. The GLCM is utilized to derive textural properties such as contrast, dissimilarity, homogeneity, energy, and correlation, which are critical in differentiating between various types of skin lesions. On the other hand, the WLD provides a robust statistical analysis of the images, enabling us to extract a compact yet informative representation of the dermatoscopic features.

Our feature extraction strategy did not merely rely on these advanced techniques but was also adaptive to the challenges presented by the dataset. Given the partial availability of high-quality segmentations and varying levels of metadata completeness, the features were extracted from full images, binary segmentations, and cropped images based on the available segmentations. This multifaceted approach ensured that we could maximize the utility of each image, irrespective of its initial condition.

In summary, our feature extraction methodology was both comprehensive and flexible, allowing for the extraction of a wide array of features that are critical for accurate classification. This methodological rigor is expected to enhance the predictive performance of our classification models, setting a robust foundation for the detailed analysis and understanding of dermatoscopic images.

0.5 Classification Strategy

The classification phase of the challenge was crucial, requiring the effective translation of extracted features into diagnostic categories. Two primary classifiers were tested: Support Vector Machine (SVM) and Multilayer Perceptron (MLP). After careful consideration and comparative analysis, the MLP was selected as the primary classifier for this task. This decision was based on the MLP's proficiency in handling high-dimensional data and its flexibility in modeling complex non-linear relationships that are typical in medical imaging data. Although both models performed commendably, the MLP demonstrated slightly better performance in terms of accuracy and handling of non-linearities in the data, making it the more suitable choice for the complexities of this challenge.

0.5.1 Why the MLP Model?

The MLP was chosen due to its layered structure, which allows for learning various levels of abstraction in data representations—a vital attribute when dealing with multifaceted dermoscopic image features. Each layer can learn to recognize different patterns and features, making the MLP particularly suitable for the nuanced differentiation required among the eight skin lesion categories defined in the challenge.

0.5.2 Model Training and Optimization

The MLP model was configured with two hidden layers, which provided a good balance between model complexity and computational efficiency. The training process involved several iterations of hyperparameter tuning, where learning rate, number of neurons per layer, and activation functions were methodically optimized based on cross-validation performance metrics on the training set.

The use of Rectified Linear Unit (ReLU) activation function in hidden layers helped in mitigating the vanishing gradient problem, ensuring effective training over deep architectures. The output layer employed a softmax activation function to provide a probability distribution over the eight classes, making the model's predictions interpretable and directly usable for classification.

0.5.3 Interpretation of Results

The results from the MLP were encouraging, demonstrating robust classification accuracy across all categories. This success is attributed to the MLP's ability to effectively leverage the deep feature set extracted from the images, which included critical textural and morphological information.

The MLP's performance affirmed the efficacy of using advanced neural networks for the classification of complex medical images and underscored the importance of thoughtful feature engineering and careful model training. The success of the MLP in this challenge suggests that similar approaches could be beneficial in other medical imaging tasks, potentially improving diagnostic accuracies and patient outcomes in clinical settings.

0.6 Results and Discussion

The evaluation of the classification model on the dermoscopic image dataset highlighted the efficacy of the Multilayer Perceptron (MLP) model. The model achieved a commendable public score of approximately 0.69 and a private score of approximately 0.678. These scores reflect the model's strong predictive performance and its generalizability to unseen data, which is critical in medical diagnostic applications.

0.6.1 Model Strengths and Limitations

These results validate the effectiveness of the feature extraction and preprocessing strategies employed before model training. The MLP's ability to integrate and interpret the comprehensive feature set extracted from the images played a crucial role in achieving these scores. However, the results also suggest areas for potential improvement, particularly in enhancing the model's ability to handle data irregularities and rare cases more effectively.

0.6.2 Future Improvements

Future efforts could focus on refining the feature extraction process or exploring more sophisticated neural network architectures that could capture complex patterns in data more effectively. Additionally, expanding the dataset and incorporating more varied examples could help improve the model's accuracy and robustness, particularly on the private score metric.

Overall, the MLP model's performance in this challenge offers promising prospects for the use of machine learning in dermatoscopic image classification and sets a benchmark for future research in this vital area of medical diagnostics.

References

1. IEEE Xplore Digital Library. Available online: http://ieeexplore.ieee.org/document/918473/
2. ScienceDirect. "Title of the Article," Journal Name, Volume, Year. Available online: https://www.sciencedirect.com/science/article/pii/S0933365712001108#bib0180
3. ScienceDirect. "Title of the Article," Journal Name, Volume, Year. Available online: https://www.sciencedirect.com/science/article/pii/S0957417416302354#bib0023
4. ScienceDirect. "Title of the Article," Journal Name, Volume, Year. Available online: https://www.sciencedirect.com/science/article/pii/S0933365713001589
5. Nature. "Title of the Article," Nature, Volume, Year. Available online: https://www.nature.com/articles/nature21056
6. HAL. "Title of the Article," University, Year. Available online: https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01250955/document
7. ISIC Workshop 2021. Available online: https://workshop2021.isic-archive.com/
8. ISIC Workshop 2020. Available online: https://workshop2020.isic-archive.com/#paper