IMA205 - Challenge
Dermoscopic Image Classification

Carlos Gruss
Télécom Paris

April 2024

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. Its incidence has increased during the last decades, especially in the areas mostly populated by white people [5].

The most effective treatment is an early detection followed by surgical excision. This is why several approaches for skin cancer detection have been proposed in the last years (non-invasive computer-aided diagnosis (CAD)) [11].

The goal of this challenge is to classify dermoscopic images of skin lesions among 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 first approach was to manually extract what are known as *ABCD features* for skin lesion diagnosis [4] [13], which refers to the asymmetry, border structure, color, and dermatoscopical characteristics that can be extracted from an image and its segmentation, and mirrors the process a doctor would go through when manually classifying a skin lesion. This extracted features were combined with the available metadata of the patients in the dataset and fed through a classifier. In this case, a Support Vector Machine (SVM) and a simple Multi-Layer Perceptron (MLP) were tested.

The second approach consisted in extracting deep features from the images themselves using a simple Convolutional Neural Network (CNN). This approach aimed to capture more nuance from the available images than is possible with simple manually extracted features. Deep learning methods have shown incredible performance in skin lesion classification, using state of the art architectures such as ResNet [6].

DATASET

The ISIC (International Skin Imaging Collaboration) skin lesion dataset, designed for the development and evaluation of machine learning models, is widely used in dermatology research and computational diagnosis. This publicly accessible dataset is particularly valuable for diagnosing skin lesions [2] [3] [19].

Challenges posed by the dataset

An undisclosed subset of this public dataset was made available for the challenge, presenting several challenges for the intended analysis.

Firstly, the dataset comprised only 18,998 labeled observations. Although this number is significant, it is relatively limited for contemporary image classification tasks. To mitigate this, transfer learning techniques could be utilized, such as employing deep learning architectures pre-trained on extensive datasets like ImageNet [10]. However, in the spirit of the challenge's rules, which restricted us to use only the provided data, it was decided to not pursue this approach.

Secondly, ground-truth segmentations were available for only approximately 10% of the data. This scarcity hindered the manual extraction of ABCD features from the images, a method dependent on precise segmentations. Additionally, metadata for about 2,000 patients was missing from the training dataset. Previous studies have demonstrated significant correlations between the location, age, and gender of patients and types of skin lesions [16] [18]. The absence of this metadata compromises the dataset's discriminative power when employing a feature-based classification approach.

Finally, the dataset's highly unbalanced nature presented the most significant challenge, as illustrated in Figure 1.

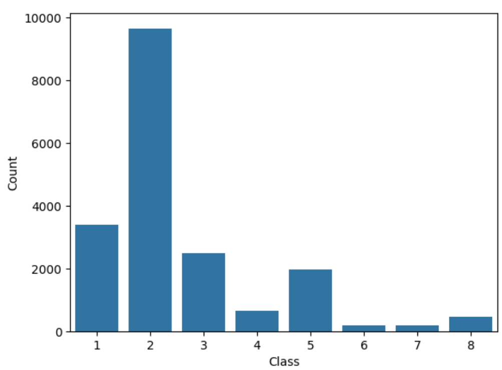

Figure 1: Distribution of classes in the dataset.

Proposed solutions

To address the limited sample size and the dataset's unbalanced nature, data augmentation techniques were applied. Specifically, the adaptive synthetic sampling approach (ADASYN) was employed [9]. Figure 2 illustrates the class distribution after the pre-processed training dataset underwent augmentation.

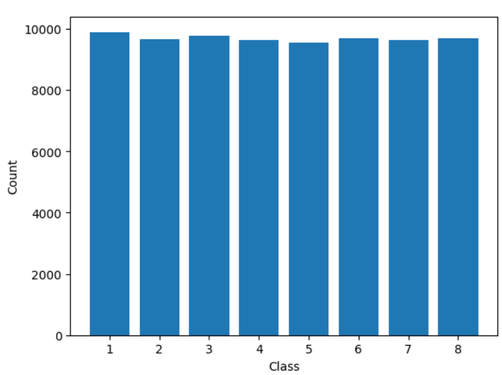

Figure 2: Distribution of classes after ADASYN.

The implementation of this strategy resulted in an approximate 10% improvement in performance metrics. However, the data augmentation was only performed on the feature-only classification approach, and not when using the images themselves as inputs. Data augmentation techniques also exist when using the images as input [14], but they weren't explored due to lack of time and computational resource availability.

Regarding the missing metadata, several strategies were considered. Eliminating records with missing data was initially contemplated but was dismissed due to the substantial amount of lost information relative to the total dataset size. Instead, the median age from the training dataset was used to impute missing age data. For missing categorical data such as class and position, a placeholder value of "unknown" was introduced. These features were then one-hot encoded, and the columns corresponding to the "unknown" values were subsequently removed. This method aims to acknowledge the absence of data without introducing noise through arbitrary substitutions.

A deep learning approach was chosen in order to obtain the missing segmentations, which is discussed in the following section.

SEGMENTATION

To address the issue of missing segmentations, a deep learning architecture based on ResNet, named MFSNet [1], was adapted and implemented for our dataset. The model was trained using all available ground-truth segmentations. Subsequently, MFSNet was employed to predict the missing segmentations. The original article also proposes an in-painting script which was used as a pre-processing to the segmentation, which helps reduce the effect of hair (or other elements that may occlude the lesion) on the segmentation.

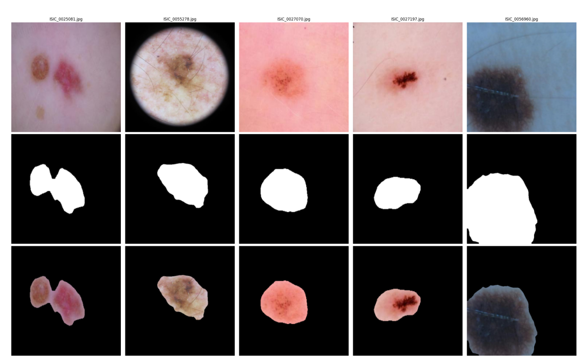

Figure 3: Segmenations obtained with MFSNet.

Prior to this application, the segmentation method's effectiveness was evaluated by dividing the existing ground-truth segmentations into training and testing sets. The DICE coefficient was calculated for the segmentations produced by MFSNet, yielding an impressive score of approximately 90%. This step was also used in determining the optimal threshold for the outputted MFSNet fuzzy segmentations: The DICE score was computed for a series of candidate thresholds and the one maximizing the score was chosen.

Despite being more resource-intensive and slower than other segmentation methods, such as OTSU thresholding, the MFSNet approach was selected due to its superior segmentation quality. High-quality segmentations are crucial for achieving accurate results in the feature-based classification approach.

FEATURE EXTRACTION

As previously discussed, the manual extraction of features using the ABCD approach was pursued. The asymmetry, border, and color features were selected based on the comprehensive analysis provided by Ganster et al. [5]. Additionally, to capture dermatoscopical characteristics, texture information was extracted from the images. This was achieved using the Gray Level Co-occurrence Matrix (GLCM) and the Weber Local Descriptor (WLD), inspired by the approach of Reddy et al. [15].

The GLCM can be employed to derive texture properties [7], while we can extract statistical properties from the WLD to obtain a compact representation of the descriptor [8]. A full description of the features implemented can be found in Table 1.

Note that some of the features are computed based on the full images, some on the extracted binary segmentation, and some on the image after cropping it using the segmentation as a mask.

Furthermore, observe that the number of features that can be manually extracted with this approach is arbitrary and fine-tunable. For example, it would be possible to extract more statistical features from the border and color categories by computing histograms, or to tune the amount of texture features by modifying the windows used when computing the GLCM, or the amount of bins in the WLD descriptor histogram. This is a difficult thing to optimize, but there's potential to improve the results obtained by further adjusting these parameters.

Table 1: Summary of Image Features Extracted

Feature Category	Features
Assymetry Features (6)	Area, Perimeter, Circularity, Bulkiness, Solidity, Eccentricity
Border Features (8)	Mean and std of the gradient in the border of the segmentation for each RGB channel
Color Features (6)	Mean and std of the RGB channel of the cropped image
Texture Features (156)	WLD Mean, std and histogram; GLCM properties (Contrast, Dissimilarity, Homogeneity, Energy, Correlation, ASM)

The amount of texture features relative to the other features reflects the belief that these features will be more robust to potentially bad segmentations and brightness variations.

CLASSIFICATION

Two distinct classification approaches were explored in this challenge. The first approach relied on manually extracted features, while the second combined these features with deep learning-derived features.

For the manually extracted features, an SVM and a simple MLP were trained. The best parameters for the SVM were determined through several iterations of Grid Search Cross-Validation, and can be found in Table 2.

Parameter	Value
Kernel	rbf
C	100
Gamma	0.01

Table 2: Best parameters found for the SVM

For the MLP, various architectures were tested, ultimately finding that a model with two hidden layers performed best. This area presents numerous design challenges and opportunities for optimization, significant improvements could likely be achieved with additional experimentation. The best performing MLP model's architecture can be observed in Figure 4.

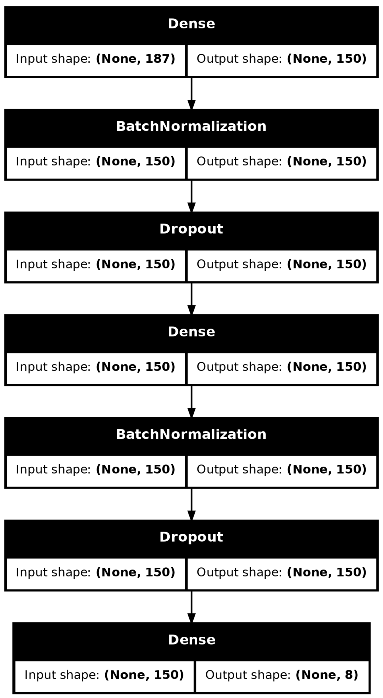

Figure 4: MLP model architecture.

Both models, the SVM and the MLP, yielded comparable results. However, the MLP was particularly noted for its faster training times, greater tunability, and higher potential for performance enhancement.

Deep learning for skin lesion classification is a very popular field, with many advanced Convolutional Neural Network (CNN) architectures available [17]. For this challenge, a simple CNN architecture was implemented and integrated into a multi-input neural network alongside the manually extracted features, aiming to enhance performance. The architecture proposed can be seen in Figure 5.

This approach was chosen because a simple CNN alone would not fully capture the dataset's complexity, and more advanced architectures like ResNet were considered too resource-intensive given the time and computational constraints of this challenge. The proposed architecture benefits from its efficiency and its ability to leverage the already computed manual features, which had proven effective at capturing significant structural

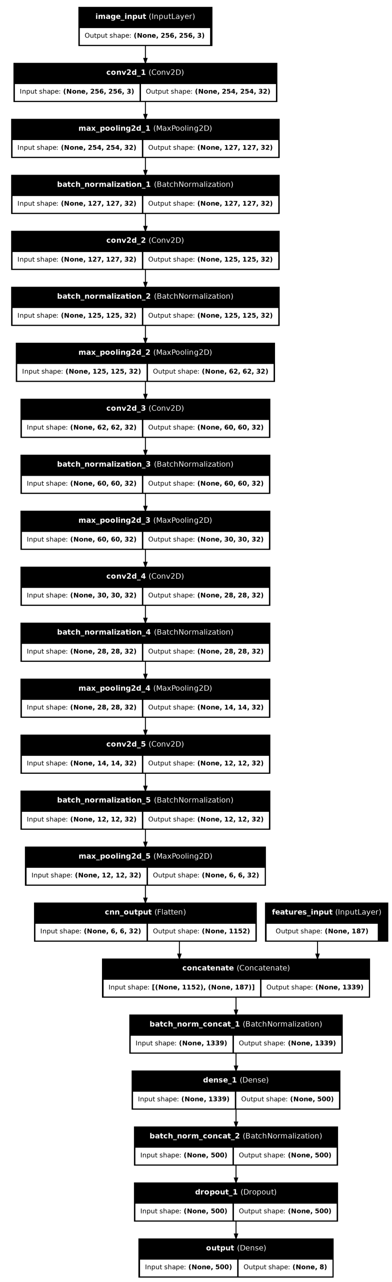

Figure 5: Multiple input model architecture.

aspects of the dataset. The strategy was to train a feed-forward network that combines manually computed features with deep-learning features from a CNN, enhancing the model's ability to discern further nuances within the dataset.

CONCLUSION

Results

The results of the Kaggle challenge for each method are summarized in Table 3.

Method	Public Score	Private Score
SVM	0.577	0.515
MLP	0.582	0.522
CNN	0.552	0.525

Table 3: Comparison of method scores on public and private datasets

The data indicate that the Convolutional Neural Network (CNN) exhibited the lowest performance, despite having the highest computational complexity, leading us to conclude that this approach was inefficient for the task at hand. While multi-input/multi-output networks can be effective in certain scenarios [12], this architecture would require further refinement and experimentation to justify its application here. Alternatively, employing a well-established standalone CNN architecture, such as ResNet, which has demonstrated effectiveness in skin lesion classification [6], might have been more prudent.

Comparatively, the Multilayer Perceptron (MLP) and Support Vector Machine (SVM) methods yielded closely competitive results, with the MLP slightly outperforming the SVM. The MLP also benefits from faster training and prediction times and offers greater flexibility in tuning, making it the more suitable choice among the methods evaluated for this challenge.'

The code for everything exposed here, including the modified MFSNet and the trained models can be found can be found here.

Future improvements

As previously noted, the extraction of ABCD features offers room for enhancement. There is potential to expand upon the current feature set to increase discriminative capability. Additionally, Ganster et al. have suggested more complex features that were not implemented in this challenge.

Regarding the models employed, the MLP exhibits considerable potential for improvement. It has numerous tunable hyper-parameters, and its training process is relatively fast, even on non-specialized hardware.

The multiple input model also presents significant opportunities for enhancement, though its complexity and slower training times pose challenges for fine-tuning. Implementing data augmentation techniques could potentially yield performance gains for this model, capitalizing on its architecture to better generalize across the dataset.

REFERENCES

[1] Hritam Basak, Rohit Kundu, and Ram Sarkar. Mfsnet: A multi focus segmentation network for skin lesion segmentation. *Pattern Recognition*, page 108673, 2022.

[2] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), 2017.

[3] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, and Josep Malvehy. Bcn20000: Dermoscopic lesions in the wild, 2019.

[4] RJ Friedman and DS Rigel. The clinical features of malignant melanoma. *Dermatologic Clinics*, 3(2):271–283, 1985.

[5] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20(3):233–239, 2001.

[6] Niharika Gouda and J Amudha. Skin cancer classification using resnet. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pages 536–541, 2020.

[7] M. Hall-Beyer. Glcm texture: A tutorial v. 1.0 through 3.0. The GLCM Tutorial Home Page, 2007.

[8] XH Han, YW Chen, and G Xu. High-order statistics of weber local descriptors for image representation. *IEEE Transactions on Cybernetics*, 45(6):1180–1193, 2015.

[9] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1322–1328, 2008.

[10] Minyoung Huh, Pulkit Agrawal, and Alexei A. Efros. What makes imagenet good for transfer learning?, 2016.

[11] Konstantin Korotkov and Rafael Garcia. Computerized analysis of pigmented skin lesions: A review. *Artificial Intelligence in Medicine*, 56(2):69–90, 2012.

[12] Nicolas Menet, Michael Hersche, Geethan Karunaratne, Luca Benini, Abu Sebastian, and Abbas Rahimi. Mimonets: Multiple-input-multiple-output neural networks exploiting computation in superposition, 2023.

[13] Franz Nachbar, Wilhelm Stolz, Tanja Merkle, Armand B. Cognetta, Thomas Vogt, Michael Landthaler, Peter Bilek, Otto Braun-Falco, and Gerd Plewig. The abcd rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, 30(4):551–559, 1994.

[14] Fábio Perez, Cristina Vasconcelos, Sandra Avila, and Eduardo Valle. *Data Augmentation for Skin Lesion Analysis*, page 303–311. Springer International Publishing, 2018.

[15] Dasari Anantha Reddy, Swarup Roy, Sanjay Kumar, and Rakesh Tripathi. A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification. *Procedia Computer Science*, 218:274–282, 2023.

[16] S. Ribero, L. S. Stucci, E. Marra, et al. Effect of age on melanoma risk, prognosis and treatment response. *Acta Dermato-Venereologica*, 98(7):624–629, 2018.

[17] B. Shetty, R. Fernandes, A. P. Rodrigues, et al. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, 12:18134, 2022.

[18] K. Stanienda-Sokół, N. Salwowska, M. Sławińska, K. Wicherska-Pawłowska, A. Lorenc, D. Wcisło-Dziadecka, J. Wydmański, and W. Majewski. Primary locations of malignant melanoma lesions depending on patients' gender and age. *Asian Pacific Journal of Cancer Prevention*, 18(11):3081–3086, 2017.

[19] P. Tschandl, C. Rosendahl, and H. Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, 5:180161, 2018.