

Medical Imaging - Skin Lesion Classification

IT University of Copenhagen

Matus Stanko **Lucie Navratilova** **David Poda**

matus@itu.dk luna@itu.dk dpod@itu.dk

Jan Nahalka **Oliver Souc**
jnah@itu.dk osou@itu.dk

Abstract

This project focuses on developing a classifier to identify malignant skin lesions. We manually segment images to isolate the lesions, then use algorithms to automatically extract features such as asymmetry, compactness, color variability and more to train a classifier. The effectiveness of the classifier is evaluated by its ability to accurately detect malignant lesions. Our goal is to explore methods to enhance diagnostic tools and aid the detection of dangerous skin lesions.

1 Introduction

Skin diseases and skin cancer are more common than many of us realize. Take melanoma, for example, the most deadly type of skin cancer, which affects hundreds of thousands of people worldwide. (Cancer.Net Editorial Board, 2023) Spotting a dangerous skin lesion is not easy for the untrained eye. It might look like just a regular mole or a harmless rash, leading many to ignore these signs until it is potentially too late.

Technology has advanced exponentially in the recent years, which leads us to the question - could we find a way to accurately analyze skin lesions and identify key features that would aid in the treatment and prevention of skin diseases and skin cancer? After some research, we found many studies that focused on the same goal, using similar tools. This shows a trend in using advanced technologies to improve diagnostic precision, which supports our project's aim. It would not only assist dermatologists in making more accurate diagnoses,

but could also help individuals worried about new or changing spots on their skin, possibly in the form of a phone app, that would help determine whether a professional check is necessary. Early detection is crucial for successful skin cancer treatment. Having a tool that makes self-checks easier for everyone could significantly aid in preventing skin cancer.

The end goal of this project is to develop a classifier that can accurately recognize a malignant lesion from an image. We will achieve this by manually segmenting the images to obtain masks and use those to develop methods to automatically measure the lesion features, such as information about its shape like asymmetry and compactness, information about the color variability, and other features commonly found in malignant lesions, like the blue-white veil. The extracted features will be used in the classifier training process. In the end, we will evaluate the performance of our classifiers and discuss the results along with some of the limitations we encountered during the project. This reflection on our work will help us see potential improvements and understand this topic further.

1.1 Related work

A review of numerous studies about skin cancer detection and classification using neural network algorithms The article (Hermosilla et al., 2024) investigates how modern computer technologies, particularly neural networks, are being used to improve the early detection and classification of skin cancer through images. This review analyzes 45 different studies, focusing on how effective these technologies are. It details the accuracy of various methods and discusses the limitations highlighted in these studies, such as

image quality and how results are interpreted by medical experts. Specific strengths noted include high accuracy rates in some models and advancements in detection processes. However, it also points out problems like the need for high-quality data to train the models effectively and the importance of having experts check the results to make sure they are accurate. Convolutional Neural Networks (CNNs) are specifically highlighted for their effectiveness in handling image data. The review also addresses the challenges doctors face in accurately diagnosing skin cancer and how these new technologies could make diagnoses faster and more reliable. Towards the end, the article suggests improvements for these technologies and potential new areas for future research, such as enhancing algorithm robustness across diverse image conditions and better integrating clinical expertise. Although this article does not introduce a new method for classifying skin lesions, it discusses the capabilities and limitations of existing studies. This is crucial for anyone beginning research in this area, as understanding these aspects can ease the process.

Skin cancer detection using deep learning techniques The article (Dildar et al., 2021) examines the use of deep learning techniques in the early detection of skin cancer, presenting it as an advancement over traditional manual methods such as biopsies, which can be painful and slow, unlike computer based techniques, which allow for a quick analysis of skin lesions. The paper describes the methodologies for feature extraction and evaluates the effectiveness of various classifiers. It emphasizes the importance of image preprocessing to enhance visual clarity, followed by feature extraction that uses both manual and automated techniques. These features are extracted following the ABCD rules, meaning asymmetry, border irregularity, color variation, and diameter. They are identified using techniques like edge detection, color segmentation, and texture analysis.

The extracted data is utilized to train deep learning models, such as Convolutional Neural Networks (CNNs) and Artificial Neural Net-

works (ANNs). These models were trained using multiple datasets, enabling them to learn and distinguish patterns associated with both malignant and benign lesions. The study mentions better performance of CNNs compared to other algorithms. The classifiers have demonstrated high accuracy in differentiating between cancerous and non-cancerous lesions, showing the potential of deep learning to significantly improve the diagnostic processes in dermatology.

Melanoma skin cancer detection based on image processing The article (Zghal and Derbel, 2020) describes an approach to skin cancer detection using image processing techniques focusing specifically on melanoma, which is the deadliest form of skin cancer. Their technique involved a multi-stage process starting with image acquisition followed by preprocessing, which includes filtering, morphological operations, and contrast enhancement to prepare the images for further analysis. This was followed by segmentation of the skin lesions and feature extraction based on the ABCD rule, which evaluates Asymmetry, Border irregularity, Color diversity, and Diameter. These features were quantitatively analyzed and assigned different weights to calculate a Total Dermatoscopy Value (TDV). This TDV categorized each lesion as benign, suspicious, or malignant. The classifier was trained on a dataset of preprocessed images, each labeled with its corresponding TDV value. The system effectively learned to associate TDV scores with the correct diagnostic categories. The model's accuracy was then validated with a separate set of images to ensure it can reliably identify melanoma in new cases. The system demonstrated a high accuracy rate of 90%, indicating its potential as a reliable tool for dermatologists to quickly identify potentially dangerous skin lesions.

2 Data

The PAD-UFES-20 dataset (Pacheco et al., 2020) consists of 2298 images of 1641 skin lesions from 1373 patients. Each entry in the data set has a number of features (*excluding IDs*):

Table 1: Lesion Information

Lesion Details	
- Location of the lesion	
- Measurements of the lesion	
- Diagnosis	
- Symptoms associated with the lesion	
- Changes in the lesion	
- Biopsy status	

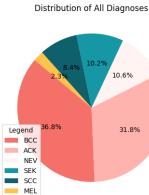


Figure 2: Distribution of diagnoses

Table 2: Patient Information

Patient Background	
- Age	
- Gender	
- Cancer history (skin)	
- Smoking status	
- Alcohol consumption	
- Lifestyle information	
- Parental background	

As with any data analysis project, it is always a good idea to take a brief look at the dataset to know what we are working with. To do this, we used pandas to explore the csv file and matplotlib to plot some visualizations of the data.

First, we wanted to know the size of our dataset, mainly to know how many patients there is. Afterwards, we explored all the different columns of the dataset to see what information it provides about the lesions. This gave us a basic idea of the data we have available.

```
image_count = len(data['img_id'])
patient_count = len(data['patient_id'].value_counts())
lesion_count = len(data['lesion_id'].value_counts())
print(f'There are {image_count} images of {lesion_count} from {patient_count} patients.')
✓ 0.0s

There are 2298 images of 1641 from 1373 patients.

print(data.columns)
✓ 0.0s

Index(['patient_id', 'lesion_id', 'smoke', 'drink', 'background_father',
       'background_mother', 'age', 'pesticide', 'gender',
       'skin_cancer_history', 'cancer_history', 'has_piped_water',
       'has_sewage_system', 'fitspatrick', 'region', 'diameter_1',
       'diameter_2', 'diagnostic', 'itch', 'grew', 'hurt', 'changed', 'bleed',
       'elevated', 'img_id', 'biopsied'],
      dtype='object')
```

Figure 1: Basic overview of the dataset

Our goal is to develop a classifier that correctly identifies whether a lesion is malignant. Therefore, we thought it might be interesting to look at the distribution of the different diagnoses in the dataset to see if they are roughly equally represented.

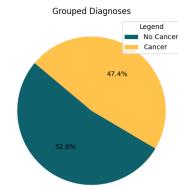


Figure 3: Distribution of grouped diagnoses

We can see that if we group the diagnoses by cancerous and not cancerous, they are split pretty evenly, with the most of our images being BCC (2.1.1) and ACK (2.1.2).

Let's take a further look at what the dataset contains. As we've seen from the pie chart above, there are numerous samples of 6 different types of skin lesions - 3 of them being skin cancer (malignant), and 3 of them being skin diseases (benign).

Table 3: Types of Skin Cancers in our dataset

Skin Cancers	
- Basal Cell Carcinoma (BCC)	
- Squamous Cell Carcinoma (SCC)	
+ Bowen's disease (BOD), clustered with SCC	
- Melanoma (MEL)	

Table 4: Types of Skin Diseases in our dataset

Skin Diseases	
- Seborrheic Keratosis (SEK)	
- Actinic Keratosis (ACK)	
- Nevus (NEV)	

2.1 Skin cancers & diseases

We will briefly describe each of the diagnoses found in the dataset we are working with, as

many of these conditions might be unfamiliar to most people. The purpose of this is to describe how these conditions manifest on the skin and their varying levels of severity.

2.1.1 Skin cancers

Melanoma (MEL) Melanoma is the most dangerous type of skin cancer. It metastasizes rapidly throughout the body, and if not treated at an early stage, it is likely to be fatal. Only 14% of patients with metastatic melanoma survive for five years. (Miller and Mihm Jr, 2006) This cancer typically presents as a change in an existing mole or the appearance of a new spot. These changes can include variations in color, shape, size, elevation, and the presence of itching or bleeding. (Cancer Council Australia, 2023a) The most significant risk factor for developing melanoma is UV exposure, especially from sunburn, along with a family history of melanoma. The most effective prevention against melanoma is doing regular skin checks and minimize UV exposure by using sunscreen daily. Early detection of MEL is crucial as it significantly enhances the chances of successful treatment and survival.

Basal Cell Carcinoma (BCC) Basal cell carcinoma is the most common malignant tumor, making up about 70% of non-melanoma skin cancers. Fortunately, in most cases it is curable when diagnosed in the early stages, leading to a relatively low mortality rate. It is often described as the least aggressive form of skin cancer. However, patients who had BCC in the past are at an increased risk of developing more severe forms of skin cancer, such as MEL or SCC. (Lear and Smith, 1997)

The biggest risk factor associated with BCC is UV light exposure. This makes lighter skin phenotypes more susceptible to developing this type of cancer. (Crowson, 2006) It can arise anywhere on the body, including on scarred or burn wounds. (Cancer Council Australia, 2023b) BCC often has no symptoms and is characterized by its slow growth and the fact that it rarely metastasizes. It can look like a pearly lump that is shiny and pale or bright pink, which can easily go unnoticed or be mistaken for a rash or non-cancerous skin disease.

Squamous cell carcinoma (SCC) Squamous cell carcinoma is one of the most common forms of skin cancer. SCC is not as dangerous as melanoma, but it can spread to other parts of your body if not treated. Every year, people in Australia die from aggressive SCCs. (Healthdirect Australia, 2023)

SCCs can appear as scaly red patches, open sores, rough, thickened or wart-like skin, or raised growths with a central depression. At times, SCCs may crust over, itch or bleed. The lesions most commonly arise in sun-exposed areas of the body. (Skin Cancer Foundation, 2024)

SCCs can also occur in other areas of the body, including the genitals.

SCCs look different on everyone. For more images, visit our Skin Cancer Pictures page. To learn more about SCC signs, symptoms and early detection strategies, go to our SCC Warning Signs page.

2.1.2 Skin diseases

Seborrheic Keratosis (SEK) Seborrheic keratosis (SEK) appears as a common non-cancerous skin growth. These skins lesions are not pre-cancerous but they may resemble other disease-like growings that are indeed prognosticated as pre-cancerous. They appear as epidermal skin tumors that commonly become visible on skin of adult and elderly individual's bodies, so they are considered as one of the most common types of human skin tumors. (Bhutta., 2023)

Actinic Keratosis (ACK) Actinic keratosis results from long-term sun exposure, which damages the outermost skin layer. These lesions may appear red or brown with a rough, scaly texture and can be flat or slightly raised. (Mayo Clinic, 2022) Treatment options vary from topical creams to procedures like cryotherapy, laser therapy, or surgical removal, depending on the lesion's characteristics. Prevention through protective clothing, sunscreen use, and reducing sun exposure is crucial to minimize the risk of ACK progressing to skin cancer. (MedlinePlus, 2022)

Nevus (NEV) Nevus, commonly known as a mole, is a benign cluster of pigmented cells.

Most adults possess several nevi, which appear as small, dark brown spots on the skin. Although generally harmless, it's important to monitor for changes in nevi, as atypical nevi can have a higher risk of evolving into melanoma. Regular skin checks are advised to track any changes. (MedlinePlus, 2017)

3 Methods

We began by manually segmenting several images in LabelStudio to create masks for use in the feature extraction process. Additional masks provided by other groups were shared by the teachers and were also utilized.

All feature extraction methods were developed in Python, primarily using the cv2 (computer vision), skimage (image processing), and numpy (scientific computing) libraries. We also used matplotlib to plot our results to verify the accuracy of our code by comparing the visualizations with our predictions.

In this section, we will briefly describe each feature and the methods we developed for its automatic extraction.

3.1 Feature extraction

3.1.1 Asymmetry

Asymmetry of skin lesions is a critical diagnostic feature when identifying malignant skin lesions. Asymmetry indicates abnormal growth patterns and is often a sign of malignancy. Typically, benign lesions are more uniform in shape than malignant lesions, which often display significant asymmetry. Our algorithm for measuring asymmetry utilizes just the binary mask of the lesion. The process begins by locating the center of the lesion and then drawing a vertical and horizontal axis through this point. We measure symmetry by flipping the lesion mask along both the vertical and horizontal axes. We then overlap the original mask with the mask flipped vertically, and repeat this with the horizontally flipped mask. Next, we calculate the vertical and horizontal symmetry scores by measuring the overlap of the original and flipped masks, or in other words, their intersection over their union. These two scores are averaged to produce a single overall symmetry

score, where 1 indicates perfect symmetry and 0 indicates complete asymmetry.

The visualization of this algorithm is illustrated in the figure below. The top row of images displays the process of measuring vertical symmetry, while the bottom row focuses on horizontal symmetry. Starting from the left, the first image presents the original mask of the skin lesion. The second image shows the mask flipped along one of the symmetry axes, and the third image depicts the overlap of the original and the flipped mask. Each image includes the axes of symmetry displayed on top of it, making it easy to identify the center of the lesion where the axes intersect.

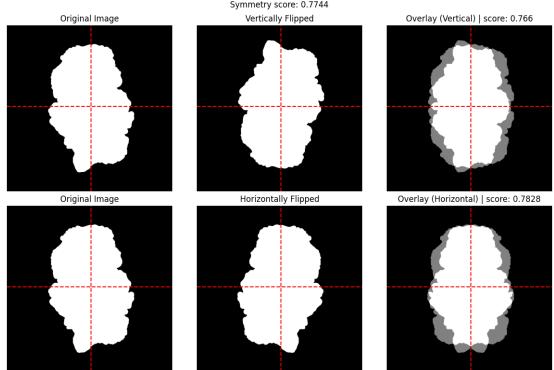


Figure 4: Asymmetry calculation visualization

3.2 Compactness

Compactness is a measure of how close a shape is to a perfect circle. This measure can help to distinguish malignant lesions, which tend to have irregular edges and a high level of asymmetry, as opposed to benign lesions, that are commonly uniform in shape. Compactness is calculated using this formula:

$$C = \frac{4\pi \times \text{Area}}{\text{Perimeter}^2}$$

where:

- **Area** is the total area of the skin lesion
- **Perimeter** is the total length of the lesion's edge

The algorithm begins by detecting the lesion edges from the binary mask provided. From

these, it identifies the largest visible shape, which is assumed to represent the lesion. Next, it calculates the area enclosed by this shape and its perimeter. These values are then plugged into the formula above to compute a compactness score, which ranges between 0 and 1. Scores closer to 1 indicate a shape that is more circular and typically benign, while lower scores suggest an irregular shape that could be indicative of malignancy.

3.2.1 Color variability

Color variability is a critical feature in recognizing a potentially cancerous lesion, as the presence of multiple colors within a lesion is often indicative of malignancy. This diversity in hues ranging from shades of brown, black, to red, and even blue or white is a significant sign of abnormality of the lesion, often associated with skin cancers like melanoma. We have come up with two approaches of analyzing color variability of the lesions in our dataset, which will be described in the following sections.

SLIC The SLIC algorithm simplifies an image into segments (also known as superpixels). It starts by placing initial points called centroids across the image. These centroids will be the core of potential segments. Each pixel in the image is then assigned to the nearest centroid based on both its color and how close it is to the centroid. After all pixels have been assigned, the centroids are moved to the average location of their respective pixels. This process is repeated until the centroids stabilize, grouping the colors in the lesion into numerous segments. The visualization of this algorithm can be seen in the picture down below:

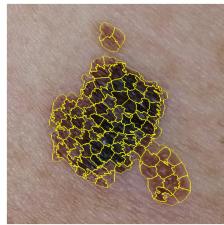


Figure 5: SLIC algorithm with 100 segments

The algorithm utilizes both the image and its

binary mask. The lesion is then isolated using the mask to ensure that the SLIC algorithm only analyzes the colors within the lesion. It segments the image and calculates the average color of each segment. These colors are then compared to a predefined list of typical colors associated with different diagnoses in our dataset using the Manhattan distance calculation to determine similarity. If a segment's average color matches the color associated with a certain diagnoses, it's noted down. The algorithm returns a proportion of the segments that match any of the top colors found in the cancerous diagnoses from all segments.

To identify the top ten colors of each diagnosis in our test dataset, we categorized the images into two groups: cancerous and non cancerous. Within each group, we analyzed the segments' color values and determined the ten most frequently occurring colors.

To provide a visual representation, we generated histograms showing the distribution of these dominant colors in the segments from both the cancerous and non-cancerous groups. These histograms help visualize the color trends and variations between the two categories.

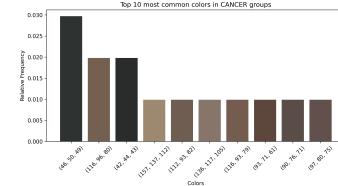


Figure 6: Top 10 most common colors in cancerous lesions

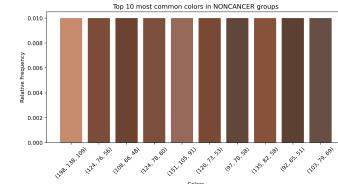


Figure 7: Top 10 most common colors in non-cancerous lesions

Color histograms Classifying lesions as either cancerous or non-cancerous relies heavily on analyzing histogram data. By extract-

ing key metrics such as mean intensity, standard deviation, and peak value from lesion images, our classifier can better understand their visual characteristics. The mean intensity gives an idea of the overall brightness of the lesion, while the standard deviation shows the variability in intensity, which could indicate malignancy-related heterogeneity. Moreover, identifying the peak intensity helps us understand the dominant features within the lesion. Using these metrics, our classifier can recognize subtle differences in lesion appearance, leading to accurate and reliable diagnoses. The code reads an image and its corresponding mask, then computes and returns statistics for each color channel (red, green, blue) within the masked region. It uses OpenCV to split the image into its respective color channels and calculates histograms for each channel using the mask. The histograms are normalized, and the code computes the mean, standard deviation, and peak value (most frequent intensity) for each color channel. These statistics are stored in a dictionary and returned.

3.2.2 Blue-white veil

The blue-white veil is a characteristic often seen in malignant lesions, especially in melanomas. It presents as an irregular, confluent patch of blue and white hues within the lesion. Recognizing a blue-white veil is critical for the early detection and treatment of cancer, as it may signal a malignant transformation within the lesion.

Our algorithm utilizes both the image and the binary mask of the lesion. It begins by isolating the lesion using the mask to ensure the detection of the blue-white veil is not affected by external elements like hair or skin markers. The image is then converted to the HSV (Hue, Saturation, Value) color space. We define a specific range of colors that we associate with the blue-white veil, set between (90, 20, 80) and (150, 150, 150). These values may be a bit hard to interpret, so we provide a short description of what each of these stand for:

- **Hue** (90-150): captures various shades of blue
- **Saturation** (20-150): detects blue shades

ranging from less saturated, which appear more gray, to more vibrant blues, though not the most intense shades

- **Value** (80-150): assists in distinguishing between lighter and darker shades of blue without including extreme whites or blacks

In our tests, which included a variety of images from our dataset and some from the internet, the main challenge was to refine the range that accurately detects the blue-white veil. This will be further discussed in subsequent sections.

We created a new mask that includes all pixels where the color falls within the defined range. We then calculate the coverage ratio by counting all the pixels where the blue-white veil was detected and dividing this by the total number of pixels in the mask. The algorithm then returns a binary decision indicating whether the blue-white veil has been detected. The visualization of this can be seen in the figure below:

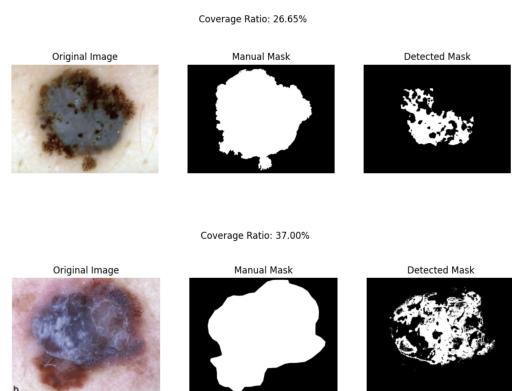


Figure 8: Blue-white veil calculation

lesion images source: (Madooei and Drew, 2013), (Popadić et al., 2017)

3.3 Classification

For our classifier training, we are focusing on two classes: cancerous and non-cancerous lesions. We chose this approach because recognizing whether a lesion is malignant is more crucial than identifying its specific type, as all cancers should be treated as soon as possible for best chance at recovery.

3.3.1 Cross-validation

We split our data into training and testing data using cross-validation with 5 folds. This method splits the data into several smaller subsets (folds) and rotates them so that each fold gets to be the test set once, with the remaining folds serving as training data. This method provides robustness against overfitting, ensuring that our classifier performs well across various datasets by maximizing the use of all our data for both training and validation. We decided to test and evaluate three classifiers: KNN(1), KNN(5), and logistic regression.

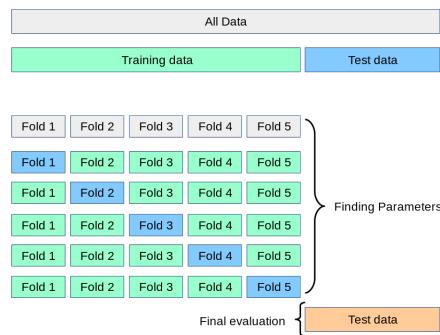


Figure 9: Cross-validation process

image source: (Scikit-learn, 2023)

3.3.2 k-Nearest Neighbors

k-Nearest Neighbors (KNN) is a straightforward machine learning algorithm. To classify a skin lesion, it calculates how far its features are from all the other features of other skin lesions in the dataset and picks the k closest ones to determine its class. KNN(1) uses just the nearest neighbor, which makes it very sensitive to specific characteristics of the nearest training example. However, this makes it less robust against random variation in the data and outliers, such as mislabeled images. KNN(5) uses five neighbors, which helps it handle outliers and random variation better, but could potentially blur the distinctions between classes. We want to see which one of these works better by comparing them to each other and see if the differences help us determine which one is more appropriate for this kind of problem.

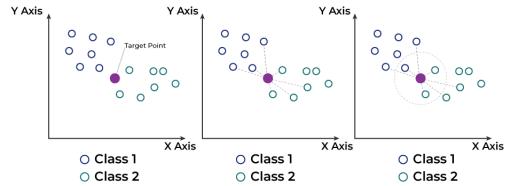


Figure 10: k-Nearest Neighbors visualization

image source: (Geeks for Geeks, 2024)

3.3.3 Logistic regression

Logistic regression is a way to model the relationship of a response binary variable and several explanatory variables. In our case, we will use this to model the probability of an input (skin lesion image) belonging to a certain class, like cancer or no cancer. The model can be described with the following equation:

$$P(Y = 1|x_1, x_2, \dots, x_n) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

where:

- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are the parameters of the model, β_0 being the intercept and $\beta_1, \beta_2, \dots, \beta_n$ the coefficients associated with the features x_1, x_2, \dots, x_n . These parameters are estimated from the data using the *LogisticRegression* function from scikit-learn, which uses parameter optimization techniques such as the Maximum Likelihood Estimator (MLE).
- x_1, x_2, \dots, x_n represent the measured features for each skin lesion.
- Y is the response variable, where

$$\begin{cases} Y = 1 & \text{means the lesion is cancerous} \\ Y = 0 & \text{means the lesion is not cancerous} \end{cases}$$
- $P(Y = 1|x_1, x_2, \dots, x_n)$ is the probability that the response variable $Y = 1$ given the measured features x_1, x_2, \dots, x_n , or in words: the probability of a lesion being cancerous given the measured features.

Each coefficient in the model quantifies the effect of a particular feature on the likelihood of choosing the positive class (cancer) over the

negative class (no cancer). Unlike KNN, logistic regression gives us the exact probability of a lesion being malignant. The decision is made by comparing the computed probability to a certain threshold, which is by default 0.5. Therefore, a lesion will be classified as malignant if the probability of it being malignant based on the model is greater than 0.5. By looking at the parameters, we can see which features were more significant when predicting malignancy. While it is already established that certain features indicate cancer (which is also the assumption we used in our project), it is beneficial to see how well our feature extraction methods capture these characteristics. This provides a great opportunity to reflect on the effectiveness of our project and lead us on the right path of improving our methods.

4 Results

Before summarizing the results of our classifier, we will briefly describe each of the evaluation metrics:

Table 5: Descriptions of evaluation metrics

Metric descriptions
- Accuracy: proportion of true results (true positives & true negatives) among the total number of cases examined
- Precision: the proportion of positive identifications that were actually correct
- Recall: the proportion of actual positives that were correctly identified
- F1 Score: balance between precision and recall, calculated as the harmonic mean of the two metrics

In our project, we believe it is important to prioritize minimizing false negatives (identifying a malignant lesion as non cancerous), while still maintaining a high level of accuracy. Therefore, we should mainly consider recall and F1 score.

The results of our classifiers are summarized in the table down below:

Since Logistic Regression models the probability as a function of the explanatory variables, which are parametrized using the training data, it relies on the trends of the entire dataset rather

Table 6: Classifier performance

Classifier	Accuracy	Precision	Recall	F1 Score
KNN(1)	0.593	0.595	0.564	0.578
KNN(5)	0.644	0.650	0.614	0.631
Logistic Regression	0.723	0.728	0.707	0.716

than individual data points like KNN. It is more robust against noise and outliers. Furthermore, it is able to handle high-dimensional data more effectively. In our case, we used 12 features as explanatory variables. This result aligns with our expectations, as Logistic Regression is a more complex model than KNN, and we expected it would give us better results. Therefore, we have decided to use Logistic Regression as our main classifier because it performed the best.

5 Open question

For our open questions, we want to discuss an interesting idea we thought about during our project.

Skin cancer can appear differently on people with various skin tones. Melanoma, SCC, and BCC are all types of skin cancer that can manifest in a variety of colors, depending on the amount of melanin in the skin. Melanin is the pigment that gives skin its color.

Individuals with higher melanin content, typically characterized by darker skin tones, have a spectrum of skin cancer manifestations that often differ from those observed in individuals with lighter skin tones. Melanoma, for instance, can present as lesions with varying shades of brown, black, or even blue in individuals with darker skin, thanks to the increased melanin pigmentation. This shows unique challenges in early detection, as these pigmented lesions may blend with the surrounding skin, potentially delaying diagnosis and treatment. On the other hand, skin cancers in individuals with lighter skin tones often manifest as lesions with hues of pink, red, or white. These contrasting colors against the lighter skin background may facilitate easier identification and prompt medical intervention. Based on this fact, potential improvement can be to train classifiers on images that have approximately similar skin color. In theory,

we for example will have 3 classifiers, each trained for different skin tone. Each of them would evaluate just those lesions for which color of patient the classifier was trained, which should increase precision to some degree.

Table 7: Colors of skin cancers on different skin tones

Skin Tone	Melanoma	SCC	BCC
Light	Pink, red, white, black, blue	Red, pink, skin-colored	White, pearly, pink
Tan	Brown, black, blue	Red, brown, skin-colored	White, pearly, pink
Dark	Brown, black, blue	Black, brown	Black, brown

6 Discussion

Based on our results, we believe we have effectively demonstrated the potential of image processing and machine learning in medical imaging. We successfully developed decent classifiers using Logistic Regression, which achieved a F1 score of 71.6%, and KNN(5), which recorded a score of 63.1%. Although our KNN(1) classifier did not achieve such great results, it still proved valuable in helping us understand which classifiers are the most suitable. We ultimately decided to use the Logistic Regression classifier, but we think all our results were surprisingly good, considering our initial expectation was to achieve scores in the 50-60% range. This assumption was due to several limitations we encountered when coding the feature extraction methods, which were beyond the scope of our project because of our limited experience and time constraints. Nonetheless, the project has taught us a lot, especially about the importance of good preparation, such as ensuring the quality of images and masks. With only four major features, it makes us wonder how much more accurate our classifier could be with additional features, or whether that might lead to overfitting. Exploring different combinations of features beyond the main (ABCD) ones could be an interesting idea for future research, especially in improving the detection of malignant lesions.

In our introduction, we proposed the possibility of developing a mobile app, given that the images were taken with a phone. Even our classifier could potentially help individuals assess whether their lesions require further medical

attention or simply encourage more people to do regular skin checks. We actually looked if such apps exist and we found quite a lot. Unfortunately, we have not had the opportunity to explore the technology behind them to see what their classifiers were trained on and which classifiers they use. This just highlights the relevance of this topic and the importance of additional research in this field to further keep enhancing classifier performance.

7 Limitations

The major limitations of our project overall were, without doubt, the quality of the images, time constraints, and our experience level. The images were taken with phones, so they were often blurry and had bad lighting. Although we tried to select both high and low quality images in the early stages of our project, we also used external images from other groups later on to train our classifiers. Manually going through those images to ensure quality would have been too time consuming, which contradicts the reason we used external images - efficiency. Better quality images or more advanced image preprocessing techniques might have improved our results. Additionally, the quality of the masks is another topic for discussion. For people without medical knowledge like us, it can be challenging to distinguish between lesions and healthy skin. We tried to skip images which we were not sure about, but it is possible that we mislabeled some by accident. The course provided a solid foundation, but as this was our first experience with image processing and machine learning, we still had so much to learn on our own. Considering this was our first project ever, we think we performed well, but there is obviously a lot of room for improvement. More experience and knowledge could enhance our feature extraction methods and thus the quality of results. Given that we only had a few months to learn everything and finish all the tasks, time constraints are reflected in our work. With more time, we would like to do our manual segmentation more carefully and select only high quality images for training our classifiers. Furthermore, exploring and implementing additional

features could also significantly enhance our results.

In this section, we will discuss the limitations of our project more in detail.

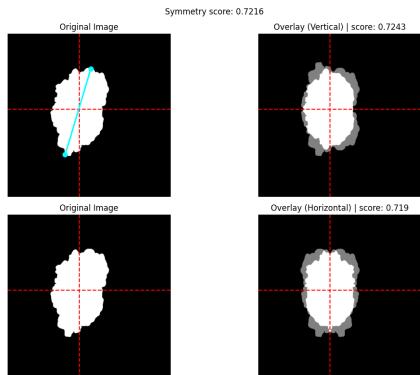


Figure 11: Not rotated

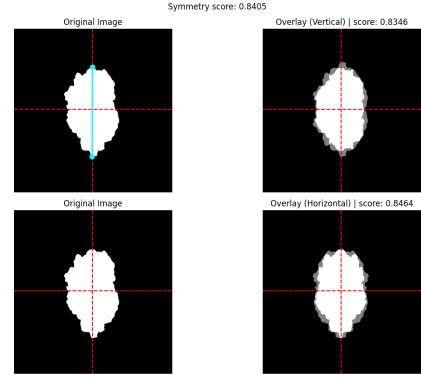


Figure 12: Rotated

Asymmetry While developing methods to measure the symmetry score, we noticed a few issues. The first problem was the uneven distribution of the lesion’s mass across the four quadrants formed by the axes, resulting in a lower symmetry score than it should be. Ideally, if we could rotate the lesion, the symmetry score would likely improve. We devised a theoretical solution: rotating the lesion so that its longest distance aligns with either the vertical or horizontal axis before calculating the symmetry scores as previously described. However, due to the complexity of the implementation and time constraints, we were unable to implement this solution. This issue, along with our theoretical approach, is illustrated in the figure below, where the second image has been manually rotated. For the rotated image, the symmetry score is higher than for the original image. The blue line within the lesion shows the longest distance, which was computed beforehand with code.

The second issue we encountered was images containing multiple skin lesions. Although an image may contain several symmetric lesions, our algorithm could return a low symmetry score because it assesses all lesions collectively. To address this, lesions should be considered individually. A potential solution is to isolate each lesion and calculate its symmetry score separately and then calculate a combined symmetry score. An example of this problem is shown in the figure below, where three reasonably symmetrical lesions receive a very low combined symmetry score.

Compactness Although the calculation of this feature is straightforward, we encountered some issues worth mentioning. To test the correctness of our algorithm, we applied it to a binary mask of a perfect circle, where the expected compactness would be 1 by definition. However, during testing, we obtained a score around 0.9. This deviation is typical with any binary mask and affects the compactness score. Digital images are composed of square-shaped pixels. This structure causes the boundaries to appear pixelated and jagged rather than perfectly smooth, which increases the measured perimeter without a proportional increase in the area, leading to slight inaccuracies in the calculation.

There are several things worth mentioning as possible solutions or improvements. One important thing to consider is the quality of the image. Higher resolution images tend to have smoother, more accurately defined edges that are less affected by pixelation. Another approach could be to refine and test various con-

tour detection parameters to determine which give the best results. However, this introduces additional considerations, such as the increased computational power required when processing a large number of images, which could be problematic when training classifiers. It may also be beneficial to explore some image preprocessing techniques, such as using Gaussian blur to smooth edges and reduce pixelation. However, the degree of blur must be carefully considered to avoid losing important details and distorting boundaries.

SLIC The SLIC algorithm was helpful for segmenting images based on the most frequent colors in the lesion, but it often mistakenly flags many segments as similar to cancerous lesions, even when the lesion is not. This happens because many noncancerous lesions have similar colors to the ones of cancerous lesions, leading to many false alarms. SLIC also gets thrown off by factors like lighting and camera quality, which can change how colors appear in images. To make SLIC more reliable for medical use, we need to ensure we’re using better and more consistent image conditions. This will help reduce incorrect classifications and improve its functionality. Due to these reasons, we have decided not to use the features extracted by our SLIC algorithm and instead rely on the other methods of analyzing color variability of skin lesions. However, if we had more time, it would be beneficial to further explore ways to improve our SLIC algorithm and potentially use the extracted features alongside other data to train our classifiers.

Blue-white veil While developing and testing our algorithm to detect the blue-white veil in skin lesions, we encountered several challenges, especially in defining the optimal range of HSV values to accurately capture the unique shades associated with the veil. One significant issue was distinguishing between the actual blue-white veil a mere light reflections on the lesion, which sometimes led our algorithm to wrongly identify these shiny areas as indicative of the blue-white veil. We considered a strategy where detection would be confirmed only if both blue hues and light (white-like)

shades were present. However, a closer look examination of the colors in images with light reflections revealed that reflective light often blends with the brown tones of the lesions, creating a spectrum of colors in the transition area, making this approach unreliable.

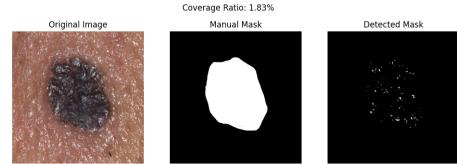


Figure 13: Light reflections detected by blue white veil algorithm

Furthermore, we were unsure about the utility of measuring the coverage ratio of the blue-white veil for our project. Given the potential inaccuracies in detection across varying image qualities, we questioned the reliability of this measurement. Deciding on a threshold for significant coverage was also challenging. We kept it regardless, because the spread or intensity of blue-white veil may signal different stages of melanoma and may change as the cancer progresses. However, due to these uncertainties, we ultimately decided on a binary output for the algorithm: it returns ‘1’ if the blue-white veil is detected and ‘0’ if not. This binary approach simplifies the decision-making process, as our classifier considers multiple features to draw conclusions.

8 Conclusion

In conclusion, the project has established a solid foundation for future research and improvement by providing us with valuable experience and highlighting several limitations that still need to be addressed. We believe we have effectively answered the initial question of whether some features of skin lesions can be reliably measured using a simple, interpretable algorithm. Achieving a 71.6% F1 score while maintaining simplicity in the feature extraction process shows the effectiveness of image processing and machine learning in medical imaging. This project highlights the immense potential not only in dermatology, but

various other medical fields. Prevention and early detection are the most effective methods for treating any health concerns, and as seen in our and other projects, machine learning has the potential to significantly transform healthcare in the future.

9 Appendix

Used programming libraries:

pandas <https://pandas.pydata.org/pandas-docs/stable/>
numpy <https://numpy.org/doc/stable/>
matplotlib <https://matplotlib.org/stable/contents.html>
opencv <https://docs.opencv.org/master/>
scikit-image <https://scikit-image.org/docs/stable/>
scikit-learn <https://scikit-learn.org/stable/documentation.html>

References

- [Bhutta.2023] Michael J. Greco; Beenish S. Bhutta. 2023. Seborrheic keratosis.
- [Cancer Council Australia2023a] Cancer Council Australia. 2023a. Melanoma — causes, symptoms & treatments. <https://www.cancer.org.au/cancer-information/types-of-cancer/melanoma>, December.
- [Cancer Council Australia2023b] Cancer Council Australia. 2023b. Non-melanoma skin cancer.
- [Cancer.Net Editorial Board2023] Cancer.Net Editorial Board. 2023. Melanoma: Statistics. <https://www.cancer.net/cancer-types/melanoma/statistics>, March. Accessed: 26th February, 2024.
- [Crowson2006] A Neil Crowson. 2006. Basal cell carcinoma: biology, morphology and clinical implications. *Modern pathology*, 19:S127–S147.
- [Dildar et al.2021] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Abdul Hakeem M Saeed, Mohammed Olaythah Al-raddadi, and Mater Hussen Mahnashi. 2021. Skin cancer detection: a review using deep learning techniques. *International journal of environmental research and public health*, 18(10):5479.
- [Geeks for Geeks2024] Geeks for Geeks. 2024. K-nearest neighbours. <https://www.geeksforgeeks.org/k-nearest-neighbours/>.
- [Healthdirect Australia2023] Healthdirect Australia. 2023. Squamous cell carcinoma.
- [Hermosilla et al.2024] Pamela Hermosilla, Ricardo Soto, Emanuel Vega, Cristian Suazo, and Jefté Ponce. 2024. Skin cancer detection and classification using neural network algorithms: A systematic review. *Diagnostics*, 14(4):454.
- [Lear and Smith1997] J T Lear and A G Smith. 1997. Basal cell carcinoma. *Postgraduate Medical Journal*, 73(863):538–542, 09.
- [Madooei and Drew2013] Ali Madooei and Mark S. Drew. 2013. A colour palette for automatic detection of blue-white veil. In *International Conference on Communications in Computing*.
- [Mayo Clinic2022] Mayo Clinic. 2022. Actinic keratosis - symptoms and causes.
- [MedlinePlus2017] MedlinePlus. 2017. Moles.
- [MedlinePlus2022] MedlinePlus. 2022. Actinic keratosis.
- [Miller and Mihm Jr2006] Arlo J Miller and Martin C Mihm Jr. 2006. Melanoma. *New England Journal of Medicine*, 355(1):51–65.
- [Pacheco et al.2020] Andre G. C. Pacheco, Gustavo R. Lima, Amanda S. Salomão, Breno Krohling, Igor P. Biral, Gabriel G. de Angelo, Fábio C. R. Alves Jr, José G. M. Essagario, Alana C. Simora, Pedro B. C. Castro, Felipe B. Rodrigues, Patricia H. L. Frasson, Renato A. Krohling, Helder Knidel, Maria C. S. Santos, Rachel B. Espírito Santo, Telma L. S. G. Macedo, Tania R. P. Canuto, and Luiz F. S. de Barros. 2020. PAD-UFES-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones.
- [Popadić et al.2017] Mirjana Popadić, Christoph Sinz, and Harald Kittler. 2017. The significance of blue color in dermatoscopy: Blue color in dermatoscopy. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*, 15, 02.
- [Scikit-learn2023] Scikit-learn. 2023. Cross-validation: evaluating estimator performance. Scikit-learn Documentation.
- [Skin Cancer Foundation2024] Skin Cancer Foundation. 2024. Squamous cell carcinoma.

[Zghal and Derbel2020] Nadia S Zghal and Nabil Derbel. 2020. Melanoma skin cancer detection based on image processing. *Current Medical Imaging*, 16(1):50–58.