A Computational Approach to Pertinent Feature Extraction for Diagnosis of Melanoma Skin Lesion

Carlos Garcia - 40220038
Sergio Vizcaino - 40291285
Adriana Danko - 40029681

Table of Contents

Introduction - What is Melanoma?

-One of the most common skin cancer types

-Early diagnostic is essential

-Can spread throughout the body

-Detection depends on geometry, color, texture, and structure of skin lesion

-Skin lesion requires key steps: preprocessing, segmentation, feature extraction, and classification.

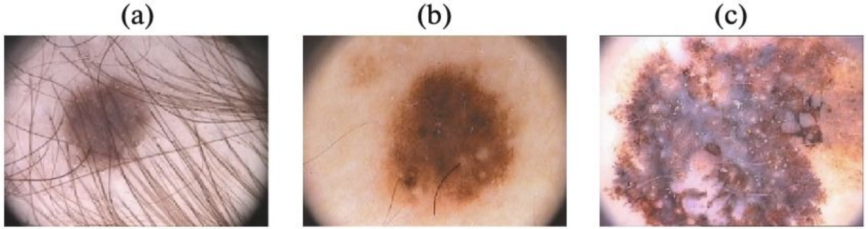


Figure 1. Sample input images: (a) common lesion, (b) suspicious lesion and (c) melanoma lesion (1)

Overview of Project

-Critical to detect as soon as possible

-Feature engineering (isolating the region of interest)

-Integration of pattern recognition algorithms and ML to enhance accuracy and efficiency of detection

-Classification - based on patterns

-Output whether or not melanoma is present or not

Dataset Description

-PH2 database

-Data from University of Porto

-Contains 200 dermoscopic images

-80 non-melanoma

-80 atypical melanoma

-40 melanoma

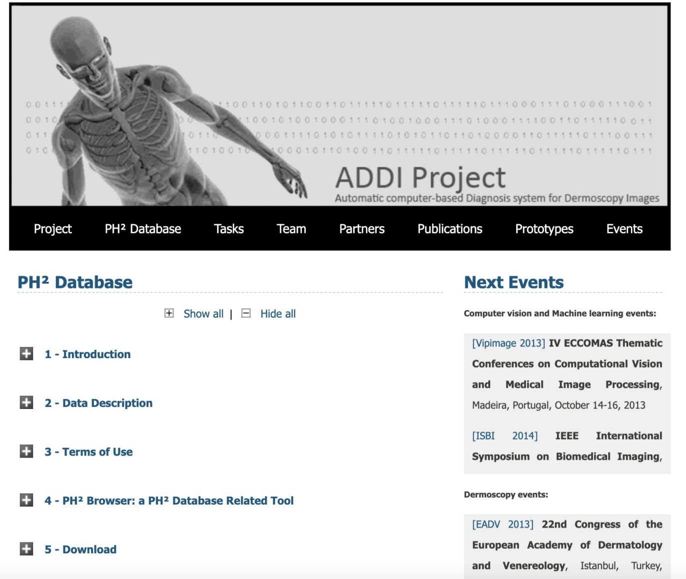
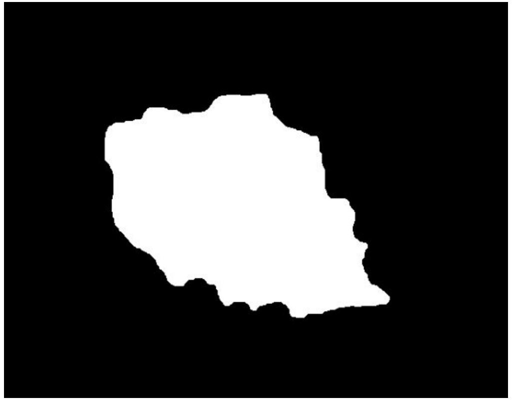


Figure 2. PH2 Database from University of Porto (2)

Feature Engineering

$$M_{ij} = \sum_x \sum_y x^i y^j f(x,y),$$

Figure 3. **Area** Formula



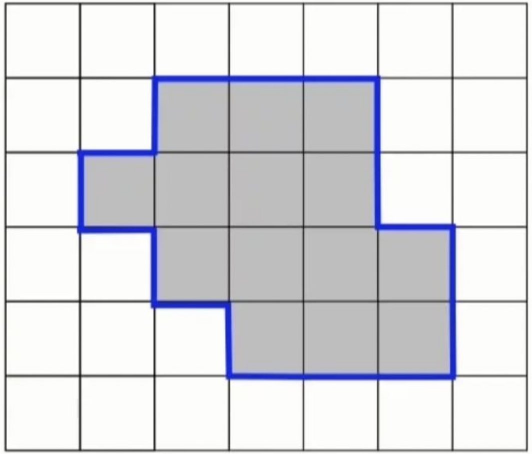
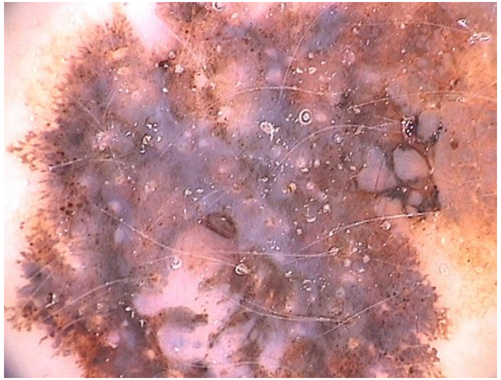


Figure 4. Determining the **Perimeter**

Feature Engineering

Colors	Red	Green	Blue
White	$R \geq 0.8$	$G \geq 0.8$	$B \geq 0.8$
Red	$R \geq 0.588$	$G < 0.2$	$B < 0.2$
Light brown	$0.588 \leq R \leq 0.94$	$0.196 < G \leq 0.588$	$0 < B < 0.392$
Dark brown	$0.243 < R < 0.56$	$0 \leq G < 0.392$	$0 < B < 0.392$
Blue-gray	$0 \leq R \leq 0.588$	$0.392 \leq G \leq 0.588$	$0.490 \leq B \leq 0.588$
Black	$R \leq 0.243$	$G \leq 0.243$	$B \leq 0.243$

Table 1. Thresholding of Red, Green and Blue channels for creating **six colors**



Feature Engineering

Area to Perimeter ratio A/P

Another measure of border irregularity is the ratio of area to perimeter of the segmented binary lesion. It is one of the early distinct signs of melanoma. Malignant lesions do have greater area to perimeter ratio than that of benign lesions.

Compactness Index

Compactness is a measure of closeness of the pixels in the shape to the center of the shape. It represents the smoothness of the lesion border. Since the most compact shape is a circle, we define compactness by **$4\pi A/P^2$** Circle has a compactness of 1, and for all other shapes, compactness is less than 1 (between 0 and 1). Melanoma lesion is considered to have ragged, uneven, blur, and irregular border and thus its compactness score deviates from one and approaches to zero.

Neural Network Architecture

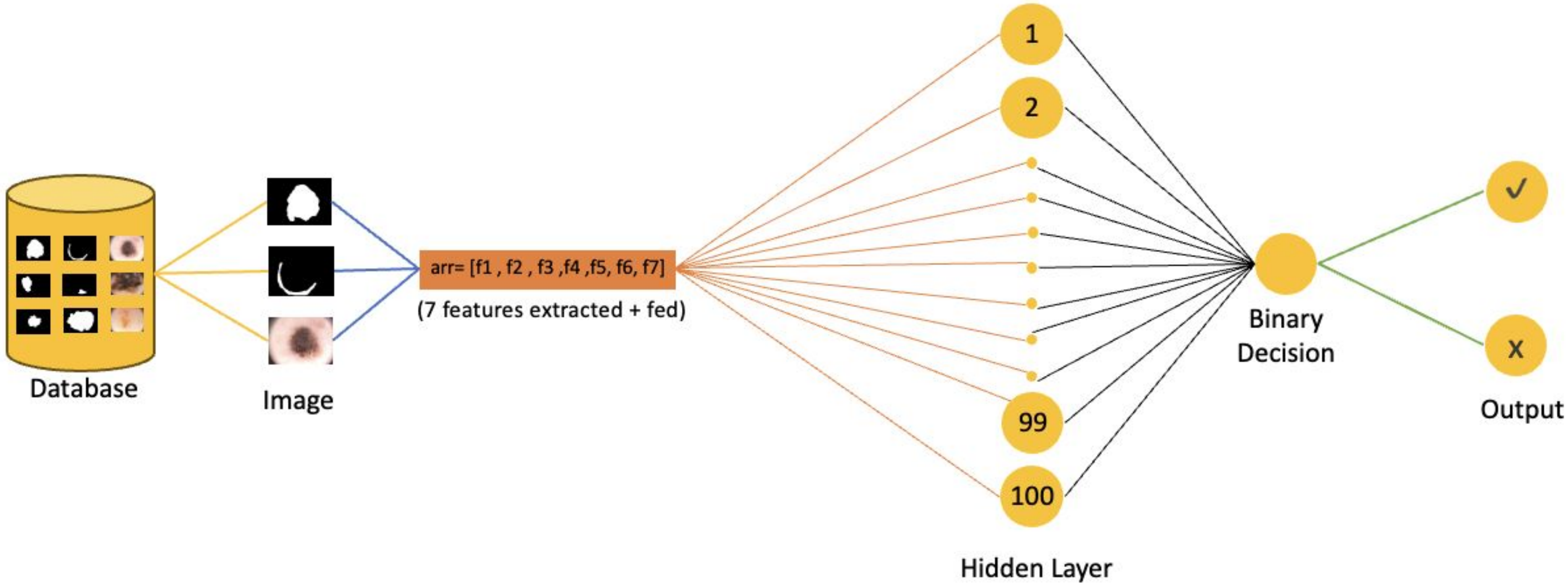


Figure 5. Overview of Neural Network Architecture

Algorithms Used for Machine Learning Training

Code

```
mlp_classifier.fit(X_train, y_train)
```

Forward Propagation

-7 features of each image is fed into the hidden layer

-NN predicts by applying weights to input data (X_train) and passing results through activation functions in the layer

Algorithms Used for Machine Learning Training

Back Propagation

-Calculates the error between the predicted output and the actual target (y_train). Adjusts weights to minimize the error by calculating the difference between predicted and actual outputs and propagating it backward through the network.

Algorithm

1. Initialize weights randomly **$\sim N(0, \sigma^2)$**
2. Loop until convergence:
3. Compute gradient, **$\partial J(W) / \partial(W)$**
4. Update weights, **$W \Longleftarrow W - \eta \, \partial J(W) / \partial(W)$**
5. Return weights

Results - Confusion Matrix

-20% of dataset used for testing

-80% of data set is used for training

-Confusion matrix

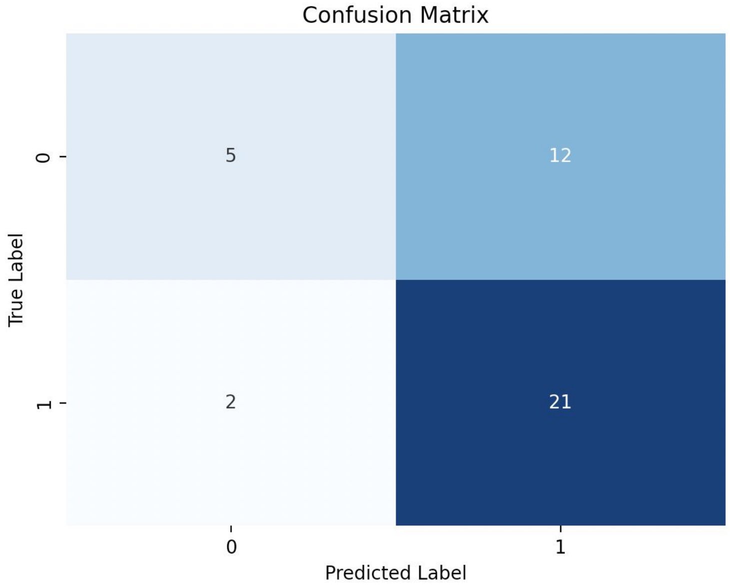


Figure 6. Confusion Matrix Results

Results - Receiving Operating Characteristic (ROC) Curve

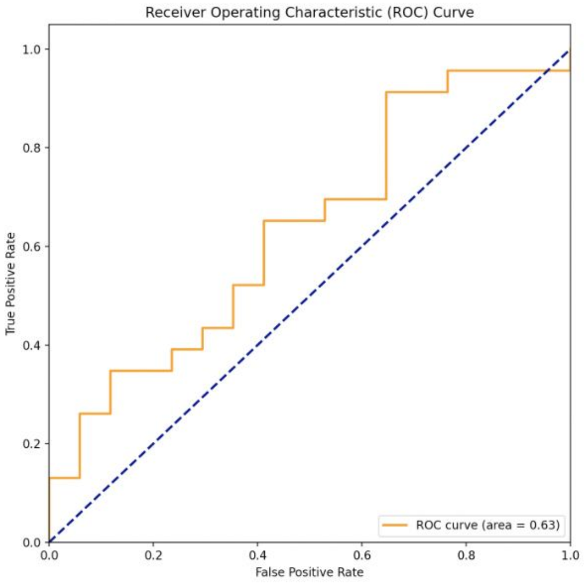


Figure 7. ROC Curve Showing Results of True Positive Rate with respect to False Positive Rate

-Showing the trade off between the true positive rate and the false positive rate

Demo

Next Steps

-Extract more features listed in the paper: best fitted ellipse, diameters…

-Finding the most optimal hyper parameters (activation function, solver, learning rate)

-Finding ways to further improve accuracy

    -Adding more neurons?

    -Have more hidden layers in the architecture?

-Further optimizing the algorithm to avoid bottleneck with one algorithm

References

(1) S.Majumdera & M.A. Ullaha. *A Computational Approach to Pertinent Feature Extraction for Diagnosis of Melanoma Skin Lesion.* (2019). Pattern Recognition and Image Analysis, Vol. 29, No. 3, pp. 503–514

(2) ADDI Project. *Automatic Computer-Based Diagnosis System for Dermoscopy Images*. https://www.fc.up.pt/addi/ph2%20database.html