Skin Cancer Classification

User Guide – How to Interpret Results

Interpreting Results (Draft)

In evaluating the performance of the Skin Cancer Classification model, several standard classification metrics are essential for understanding its effectiveness in real-world diagnostic scenarios. Since the goal of this application is to assist users in identifying different types of skin lesions from images, the performance of the model must be assessed both globally and on class-specific basis.

1. Accuracy

Accuracy represents the proportion of correctly classified samples among all predictions. While commonly used, accuracy alone can be misleading in imbalanced datasets—common in medical imaging—where one class may dominate. For instance, if benign lesions are far more common than malignant ones, a model might appear accurate simply by over-predicting the majority class.

Formula: Accuracy = (TP + TN) / (TP + TN + FP + FN)

2. Precision, Recall, and F1 Score

These metrics are particularly useful in a medical context where false negatives (missing a cancer diagnosis) are far more severe than false positives.

- Precision: Of all samples the model predicted for a specific cancer type, how many are truly diagnosed?

- Recall (Sensitivity): Of all actual cases of a specific cancer type, how many did the model correctly identify?

- F1 Score: The harmonic mean of precision and recall, balancing both metrics to give a single effective score.

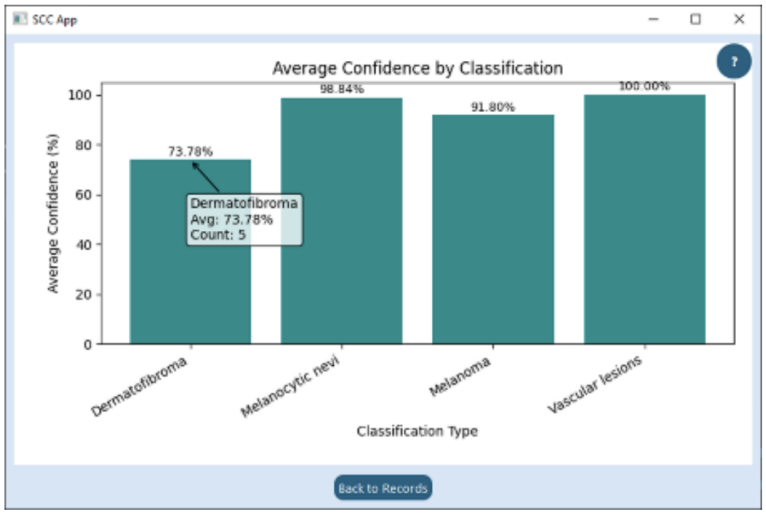
Formulas:

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 = 2 * (Precision * Recall) / (Precision + Recall)

3. Average Confidence by Classification



The “Average Confidence by Classification” histogram shows, for each lesion type, the model’s mean probability score when it predicts that class. For example, the bar for “Dermatofibroma” sits at about 73.8% with a count of 5 cases, indicating that whenever the model labeled an image as dermatofibroma, it assigned an average confidence of 73.8%. In contrast, “Melanocytic nevi” appears near 98.8%, indicating that when the model predicts nevi, the model makes the prediction with very high certainty. Similarly, “Melanoma” shows roughly 91.8% average confidence, and “Vascular lesions” is at 100%, which—depending on the sample count—suggests that those predictions were uniformly strong. A lower average confidence (like in dermatofibroma) can signal that the model finds those cases more ambiguous or that the training set contained fewer clear examples, whereas consistently high confidence implies the model learned distinct features for that class. By comparing both the heights of the bars and the sample counts (visible in the tooltip), when evaluating model predictions, we consider both the confidence level per class and the robustness of that confidence. A high confidence score based on only a few samples may be less reliable than a slightly lower score derived from a larger, more representative sample.

4.External information source in classification result page

American Cancer Society

The American Cancer Society (ACS) is a nationwide, community-based voluntary health organization dedicated to eliminating cancer as a major health problem. Founded in 1913, the ACS funds and conducts research, shares expert information, supports patients, and spreads awareness to promote cancer prevention and early detection. With a long history of advocacy and education, the organization plays a vital role in improving cancer outcomes through public health initiatives and support services for patients and families.

Mayo Clinic

Mayo Clinic is a world-renowned nonprofit academic medical center focused on clinical practice, education, and research. Headquartered in Rochester, Minnesota, it is consistently

ranked among the top hospitals in the United States. The Mayo Clinic provides expert care in a wide range of specialties, and its integrated, patient-centered approach to medicine is supported by cutting-edge research and innovation. It also serves as a trusted source of medical information for the public through its comprehensive online platform.