



Fair Melanoma Detection

Premium Partner: **Faculty of Electrical Engineering, Computer Science and Information Technology Osijek**

1. Challenge Description

1.1. Introduction and Motivation

Melanoma is one of the most aggressive forms of skin cancer, and its **early detection** is critical for improving patient outcomes. Dermatologists and researchers rely on advanced diagnostic tools to differentiate between malignant melanomas and benign skin lesions. Among these tools, the **dermatoscope** is a cornerstone in dermatological practice.

A dermatoscope is a handheld device that combines magnification and specialized lighting to examine the structure of skin lesions. This technique, known as dermatoscopy or dermoscopy, provides enhanced visualization of features beneath the skin's surface that are not visible to the naked eye. Dermatoscopy is crucial for **identifying suspicious lesions** and distinguishing them from benign growths, thus aiding in timely and accurate diagnoses.

In this challenge, participants will tackle **the problem of melanoma classification** using dermatoscopic images. By developing machine learning models, students will explore how technology can augment the diagnostic process. This challenge not only tests technical expertise but also encourages an understanding of ethical considerations, particularly fairness in AI systems.

1.2. Problem Description

Welcome to the **Melanoma Detection Challenge**! In this competition, you will step into the role of AI developers striving to create a model that classifies skin lesions as malignant or benign. Your mission is to design an algorithm that performs reliably across diverse skin types and minimizes biases.

Melanoma diagnosis involves assessing several visual cues, such as asymmetry, border irregularities, color variations, diameter, and evolution over time—known collectively as the ABCDE criteria. While these criteria guide clinical practice, automated systems can analyze large datasets more efficiently, potentially identifying subtle patterns that elude human observation.



LUNAEN

DATA SCIENCE

The dataset for this challenge, sourced from the **ISIC 2020 Challenge**, contains thousands of dermatoscopic images from a wide demographic range. These images are accompanied by metadata and ground truth labels that denote whether a lesion is malignant or benign.

Participants will use this dataset to develop and test their models. However, there's a catch: the train/validation/test split will be made by the organizers, and the test set is private. Additionally, **fairness metrics** will be a key aspect of evaluation, ensuring that models perform equitably across different skin tones.

AI fairness refers to the principle of designing and evaluating machine learning models to **ensure equitable performance** across diverse demographic groups. In dermatology, this is especially critical because research has shown that existing models often perform worse on images of darker skin tones. This disparity can lead to delayed or inaccurate diagnoses for individuals in these groups, highlighting the need for fairness-focused approaches.

In this challenge, solutions will be tested on a diverse dataset, and specific fairness metrics will be used to evaluate performance. These metrics will help ensure that the models deliver consistent and reliable results regardless of skin tone, promoting more inclusive and **ethical use of AI in medical applications**.

Participants will be provided with a pre-split dataset, including training and validation sets. The test set will remain private and used exclusively for final evaluation.

Key rules and guidelines include:

- **Dataset Usage:** Only the provided datasets can be used for training and validation. External datasets are not allowed. However, data augmentation and synthetic techniques to expand the dataset are permitted.
- **Fairness Metrics:** Evaluation will emphasize fairness by analyzing model performance across different skin tones. Precision, recall, and other metrics will be used, with particular attention to potential correlations between results and skin tone.
- **Reliability Over Accuracy:** Greater importance will be placed on the reliability of the model rather than just accuracy. Participants should balance precision and recall and aim to create a robust classifier.



2. Solution Package

The submitted solutions should contain three main parts as per the following:

2.1. Project Documentation in PDF Format

The project documentation should provide a comprehensive overview of the entire project. It must include the **project's goals, methodologies, processes, and significant findings**. This document aims to clearly outline the objectives and the steps taken to achieve them, offering valuable insights to evaluators and stakeholders. It should effectively convey the scope and outcomes of the project. Describe the used methods, choices, any relevant exploratory data analysis findings, etc. Describe the results as well as the shortcomings of the approach.

2.2. Technical Documentation in PDF Format

The technical documentation should deliver a detailed explanation of the solution's technical components. This can include **architectural diagrams, data flow illustrations, algorithms**, and other critical technical details. The goal of this document is to ensure a thorough understanding of the solution's functionality, allowing reviewers to assess the technical quality and performance of the project. Well-structured and precise documentation is crucial for showcasing the technical strengths of the solution.

2.3. Source Code

The submitted solution source code must meet the following minimum specifications:

- Include **all code necessary to reproduce results** on the validation dataset. This includes comprehensive documentation describing the general approach, the commands needed to train the model, and instructions for installing and using any libraries or dependencies.
- Provide a **pretrained model checkpoint** that can be used to perform inference on new data. The pretrained model should be included in the submission.
- Include a file named **validation_output.csv** with columns image_name and target (identical to the provided ground truth CSV). This file should contain predictions made on the validation dataset.
- Provide **a script** that runs inference on a folder of images and saves the results in a CSV file, as detailed in the remainder of this section.



Script API should look like `python predict.py <INPUT_FOLDER> <OUTPUT_CSV>`

Example: `python predict.py test_images test_output.csv`

- The script must call the pretrained model to generate predictions directly without requiring additional parameters or retraining steps.
- The output CSV file, `test_output.csv` should have columns `image_name` and `target` (matching the ground truth CSV provided).
- This script will be used to run inference on the private test set to calculate the final scores of the model.
- An example script and output file will be provided.

3. Dataset Description

The dataset for this competition is sourced from the **ISIC 2020 Challenge**, a benchmark dataset for melanoma classification. It consists of high-resolution dermatoscopic images that capture a wide variety of skin lesion types. The dataset has been curated from multiple international medical institutions, ensuring diverse representation of skin types, anatomical locations, and lesion characteristics.

The dataset will be provided to participants as a **downloadable ZIP file** containing the training and validation datasets. The training and validation datasets reflect the **real-world distribution** of skin tones, which may not be evenly balanced. To evaluate fairness and performance, a separate test dataset will be held out and kept private until the end of the competition. This test dataset will be more balanced with respect to skin color, enabling comprehensive evaluation of model fairness and reliability.

Each dataset consists of:

- **Images:** High-resolution JPEG files representing dermatoscopic images of skin lesions.
- **CSV File:** Accompanying metadata and ground truth labels for the training and validation datasets. The CSV file includes columns `image_name` and `target` (indicating malignancy, as in the original ISIC 2020 challenge), along with additional metadata that participants may use during development.



LUMEN
DATA SCIENCE

Participants should note the following:

- The model is required to use **only the images** as input for predictions. While additional columns in the CSV file (e.g., metadata) can be utilized for exploratory analysis and feature development during training, they cannot be part of the final model's inference pipeline.
- The ground truth for the test dataset **will not be provided** until the conclusion of the competition.

By structuring the test dataset to include a more balanced representation of skin tones, this challenge emphasizes the **importance of developing models** that are both **accurate and equitable** across diverse populations.



4. Scoring Criteria for Evaluation

The evaluation of submitted solutions will be based on **four key categories**, each contributing equally to the final score. These categories ensure a comprehensive assessment of the solution's quality, from methodology and code design to model performance and fairness. Below are the detailed descriptions of each category and their respective scoring distribution.

4.1. Methodology and Approach (25 Points)

This category evaluates the quality and reliability of the methodology used in the solution. Key considerations include:

- Alignment with **medical practice** and prior research.
- Adherence to **best practices in data science** and statistics to avoid common pitfalls such as overfitting or improper validation.
- Robustness of exploratory data analysis (**EDA**), preprocessing techniques, and feature engineering.
- Clarity and justification of assumptions, intuition, and methods used to develop the solution.

This category focuses on the process and methodology, independent of the final results. Evaluated based on the submitted project documentation PDF as well as the code.

4.2. Code Quality (25 Points)

This category assesses the quality of the submitted code, emphasizing:

- Readability and organization.
- Adherence to software engineering best practices, including **modularity and maintainability**.
- Quality and clarity of documentation, enabling easy understanding of the workflow.
- **Reproducibility of results** through the provided scripts and instructions.
- Computational efficiency and resource management.

Solutions should demonstrate clean, well-documented, and efficient code that is easy for evaluators to execute.



4.3. Model Performance (25 Points)

This category evaluates the overall performance of the model, focusing on:

- Accuracy, precision, recall, and other relevant metrics on the validation and test datasets.
- Stability of results with respect to random seeds and input variations.
- Consistency and reliability of predictions across different scenarios.

The evaluation considers both **quantitative metrics** and the **robustness** of the model's predictions.

4.4. Model Fairness (25 Points)

This category assesses the fairness of the model, ensuring equitable performance across different demographic groups. Key considerations include:

- Evaluation of specific AI fairness metrics using skin color as a
- Analysis of result discrepancies **with respect to skin color**.

The emphasis is on ensuring the model does not exhibit biases and maintains high reliability for all subsets of the population.

4.5. Scoring Table

CATEGORY	POINTS
Methodology and Approach	25
Code Quality	25
Model Performance	25
Model Fairness	25
TOTAL	100



5. Additional Resources

<https://shelf.io/blog/fairness-metrics-in-ai/>

https://fairlearn.org/v0.12/user_guide/index.html

<https://web3.arxiv.org/abs/2411.12846>

https://arxiv.org/abs/2208.10013?utm_source=chatgpt.com

<https://challenge2020.isic-archive.com/>

<https://www.kaggle.com/c/siim-isic-melanoma-classification/leaderboard>

<https://arxiv.org/abs/2202.02832>

<https://arxiv.org/abs/1703.03702>

<https://doi.org/10.1038/s41597-021-00815-z>