



IPSC IT FEST 2026

Judging Rubric

The Black Box

The Doon School

DEHRADUN

Judging Rubric

1. Total

1.1 Total points available: 100

2. Criteria

2.1 Phase 1: Autopsy & Failure Prediction — 20 points

Assessment	Descriptor
Full marks	Accurately logs all class counts, lighting issues, and background biases; correctly predicts primary model failure modes with clear justification.
Partial	Identifies some dataset flaws but misses key background/color biases or provides weak failure predictions.
Zero	Fails to complete audit sheet or provides vague, incorrect predictions.

2.2 Phase 2: Baseline Model Execution — 10 points

Assessment	Descriptor
Full marks	Correctly uploads dataset to The Black Box AI Deployed Model, logs all initial metrics, baseline accuracy, and confidence scores accurately.
Partial	Model trained, but metrics are partially missing or mislogged on submission form.
Zero	Model not trained or baseline metrics not recorded.

2.3 Phase 3: Failure Hunt & Confusion Matrix — 25 points

Assessment	Descriptor
Full marks	Evaluates all 30 test images correctly, populates complete confusion matrix, tags all errors accurately by cause (color, background, angle, etc.), and correctly identifies top 2 failure modes.
Partial	Matrix partially filled, minor misclassifications untagged, or top 2 failure modes improperly prioritized.
Zero	Confusion matrix incomplete or error causes completely misidentified.

2.4 Phase 4: Override Strategy & Post-Patch Accuracy — 30 points

Assessment	Descriptor
Full marks	High test set accuracy post-fix; patch photos (≤10) precisely target identified biases without introducing new errors; clear hypothesis and before/after comparison logged.
Partial	Moderate accuracy improvement; patch photos partially address flaws or slightly overfit to the test set.
Zero	Accuracy unchanged or degraded; exceeded 10-photo limit or took non-targeted photos.

2.5 Phase 5: Presentation & Q&A Defense — 15 points

Assessment	Descriptor
Full marks	Clear 6-minute presentation of diagnosis, strategy, and trade-offs; answers judge Q&A with deep technical understanding of data bias.
Partial	Presentation covers basics but lacks depth on trade-offs; hesitant during judge Q&A.
Zero	Unclear presentation, unable to justify data patch decisions during Q&A.

3. Who scores what

Phase 1, 2, 3, & 4 Worksheets & Accuracy: Evaluated by Judges post the Phases, also through the submissions.

Phase 5 Presentation & Defense: Evaluated by the designated Judge Panel during oral presentations.

4. Tie-breaks

4.1 In case of a tie, the following parameters will be taken into consideration in the same order:

• Highest classification accuracy achieved on the 30 hidden test images following Phase 4 (Override).

• Highest score awarded in Phase 1 (Autopsy & Failure Prediction).

• Highest score awarded in Phase 3 (Confusion Matrix & Error Tagging Accuracy).

• Lowest total time taken to submit deliverables across Phases 1 through 4. (Rulebook §3.4).