CS3: Extreme Weather Image Classification - Student Rubric

Goal: Successfully construct and rigorously evaluate a Deep Learning model for binary classification of satellite imagery (Extreme vs. Normal Weather).

Task Summary

Your work begins with the original 9,081 image dataset (Harvard Dataverse) and must proceed through the complete pipeline: data normalization, class consolidation, model development (baseline and advanced), full testing, and interpretable analysis.

Criteria for the Student Task

This section details the student's deliverable and the criteria used to evaluate their success in replicating and extending the weather classification pipeline.

Spec Category	Spec Details
Data Preparation	**Replication of Pipeline:** The student must successfully demonstrate the initial data science steps: - Consolidation of 5 original classes down to 2 (Extreme vs. Normal). - Image standardization (resizing to 224x224 and normalization). - Execution of a **stratified 80/10/10** train/validation/test split.
Model Development	**Advanced Architecture Use:** The student is required to implement and train at least one sophisticated model: - Implementation of a **Transfer Learning architecture** (e.g., MobileNetV2 or EfficientNetV2), using pre-trained weights and modifying the final classifier head. - (Exceeds requirement if a comparative **Baseline CNN** is also implemented).

Performance Metrics	**Target F1-Score Attainment:** The final model must be evaluated on the Test Set. The student achieves success by meeting or exceeding these key performance indicators: **Meets:** Achieves a Macro-Averaged F1-Score between **0.85 and 0.89**. **Exceeds:** Achieves a Macro-Averaged F1-Score of **0.90 or greater**.
Model Interpretation	**Grad-CAM Visualization:** The student must generate empirical evidence to prove the model's focus is meteorologically sound: - Generation of clear **Grad-CAM heatmaps** showing model attention for both correct and incorrect predictions. - A written analysis of the heatmaps, confirming the model attends to key features (e.g., storm eyes, dust plume centers).
Documentation	**Code Quality and Reflection:** The final submission must be professional, runnable, and reflective: - Code (Scripts/ folder) must be well-commented and **fully reproducible**. - Final Reflection (300-500 words in README.md) must substantively address the choice of Transfer Learning model and assess the trickiest analytical trade-off (e.g., Class Collapsing vs. Resizing).