Summary Report : Airflow Prediction Rate

1. Overview

I focused on predicting airflow rates using a minimal set of **four explainable features** derived from infrared (IR) thermal video. Specifically, I used:

1. **Delta T:** The measured temperature difference between outside and inside conditions.
2. **Bright Region Gradient 0–5s:** Average intensity change in the consistently bright region during seconds 0–5 of the video.
3. **Bright Region Gradient 5–10s:** The same calculation for seconds 5–10.
4. **Bright Region Gradient 10–15s:** For seconds 10–15.

These features capture **how intensities evolve** over multiple time segments in the brightest thermal region, along with the overall temperature difference.

2. Current Model Performance

Using **Leave-One-Out Cross-Validation** on all 22 video samples, I tuned and evaluated six different regressors (Linear Regression, Ridge, Lasso, SVR, Random Forest, and Gradient Boosting). The best model was **Ridge Regression** with an optimal alpha=0.1 parameter. Here is a brief overview of its results:(The model may change depending on the new features and their relationship)

- **Mean Squared Error (MSE):** 0.0486
- **Root Mean Squared Error (RMSE):** 0.2205
- **Mean Absolute Error (MAE):** 0.1815
- **R^2 (Variance Explained):** 0.4703 (~47%)

In other words, with these four features alone, the model explains **about 47% of the variance** in airflow rate across the 22 videos.
I believe that with new features extracted from the video, the accuracy would also increase.

3. Interpretation

- Each of the four features is **physically interpretable**:
  - Delta T: The ambient temperature difference influencing heat transfer.
  - Three temporal gradients: How quickly or slowly the hottest pixels shift over discrete time intervals, presumably tied to airflow movements.
- The model's moderate R^2 indicates we are capturing nearly half of the variability in airflow rates, suggesting that more nuanced features or additional data may further improve accuracy.

4. Next Steps and Further Improvements

1. **Additional Feature Engineering**
   - **Spatial Characteristics:** Incorporate region area, perimeter, or

centroid movement to account for how the hot spot's shape or position evolves.
    - **Intensity Distribution:** Explore percentile values (e.g., 25th, 75th percentile) or standard deviation within the bright region to capture heterogeneity.
    - **Refined Temporal Analysis:** Use smaller or overlapping windows (e.g., 2–3s) to see if finer granularity yields more predictive power.
2. **Feature Explainability and Documentation**
    - Each feature will be **documented** with its physical meaning to maintain transparency.
    - Simple parametric or tree-based methods will allow analysis of feature importances or coefficients, ensuring the model remains interpretable.
3. **Cross-Validation & Potential Hold-Out**
    - I will continue using **Leave-One-Out Cross-Validation** to maximize the use of the 22 videos, but may also test a small hold-out set if needed to provide an independent final check.
4. **Iterative Testing**
    - After adding or adjusting features, I will **track improvements** in MSE, RMSE, and R^2, to verify each feature's contribution.

5. Conclusion

With the current four-feature setup, the model's best R^2 is approximately **0.47** under Ridge Regression. This serves as a solid baseline result, indicating that the temporal gradients of the brightest region combined with the temperature difference do provide a meaningful, albeit partial, explanation of the airflow rate.

Moving forward, I plan to **expand the feature set** in a systematic and well-documented manner, focusing on physically interpretable metrics (spatial, intensity distribution, refined temporal windows). I anticipate that these enhancements will improve the model's predictive accuracy and clarify the thermal-flow relationship observed in the IR data.