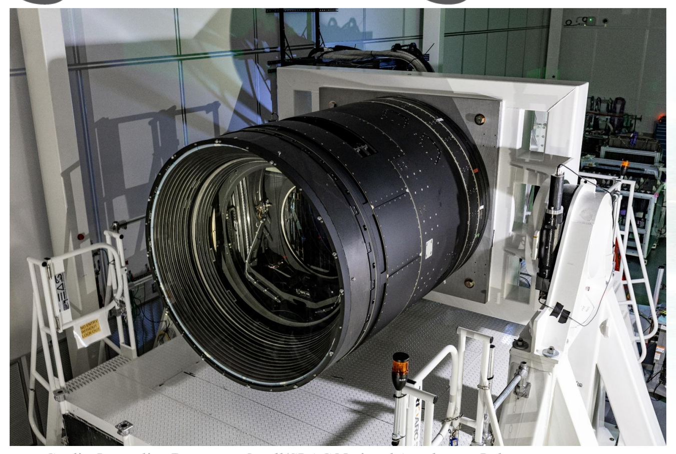
Twinkle Twinkle Little Star, is That Really What You Are?

Stanislaus State

Jonathan Rodriguez
Undergraduate
Physics and Astronomy
Advisor: Dr. Brian Morsony CSU Stanislaus

TRIO
RONALD E. MCNAIR POST-BACCALAUREATE ACHIEVEMENT PROGRAM

VERA C. RUBIN OBSERVATORY



Abstract

Accurately distinguishing stars from galaxies becomes increasingly difficult at fainter magnitudes — yet this task is critical for both cosmological measurements and studies of galaxy evolution. In this project, we analyzed approximately 750,000 simulated sources from the Rubin Observatory's Data Preview 0 (DP0), spanning AB magnitudes 23 to 29.

We evaluated the LSST pipeline's star-galaxy classifications against simulated "truth" labels across three magnitude bins: 23–25, 25–27, and 27–29. Our results revealed a sharp decline in classification accuracy, dropping from over 80% at bright magnitudes to below 50% for the faintest objects.

To address these challenges, we tested two strategies:

- **Shape-Based Filtering**: Applying cuts based on morphological features (e.g., axis ratios) allowed us to recover 10–15% of galaxies misclassified as stars.
- **Machine Learning Enhancement**: A preliminary classifier trained on ugrizy photometry and shape parameters improved classification by an additional ~15% at faint magnitudes.

These findings underscore the limitations of traditional classification methods and demonstrate the promise of advanced, feature-rich approaches for improving star–galaxy separation in deep survey data.

Why Is This Important?

Transient Science:

- Accurate object catalogs are essential for distinguishing genuine supernovae and variable events from stellar or galactic contaminants. Misclassifications can lead to false detections and missed discoveries.

Cosmological Precision:

- Incorrectly identifying faint galaxies as stars can skew measurements of large-scale structure and bias our understanding of dark matter distribution.

Galaxy Evolution:

- Reliable classification enables accurate galaxy counts and morphology analyses, which are fundamental to studying how galaxies form and evolve over cosmic time.

Maximizing Rubin's Potential:

- Robust classification ensures the scientific community fully leverages Rubin's massive datasets, transforming our view of the universe.

Acknowledgements

- Vera C. Rubin Observatory (DP0)
- McNair Scholars Program
- TRIO Student Support Services
- ASPIRE Research Initiative
- Vera C. Rubin Observatory
- CSU Stanislaus
- Dr. Brian Morsony
- Dr. Nicole Cochran
- Dr. Ellen Bell

Introduction and Background

The Vera C. Rubin Observatory is a next-generation, ground-based facility designed to explore the dynamic universe. It houses the Large Synoptic Survey Telescope (LSST) — an 8.4-meter telescope equipped with the largest digital camera ever built, boasting 3.2 gigapixels.

Over its planned 10-year mission, Rubin will repeatedly image the entire southern sky every few nights, creating an unprecedented dataset in both depth and cadence. This ambitious effort, known as the Legacy Survey of Space and Time (LSST), is expected to transform nearly every area of astrophysics.

Scientific Goals Include:

- Cataloging billions of stars and galaxies
- Mapping the structure and formation history of the Milky Way
- Monitoring transient events such as supernovae, variable stars, and near-Earth asteroids
- Probing dark matter and dark energy, the invisible scaffolding of the cosmos
- And discovering phenomena we haven't even imagined yet

First public data release: June 30th, 2025

Methodology

We analyzed approximately **750,000 simulated astronomical objects** from the **Rubin Observatory's Data Preview 0 (DP0)**. These sources span AB magnitudes **23 to 27**, representative of the faint end of Rubin's expected dataset.

To evaluate classification performance, we:

•**Divided the sources into two magnitude bins**:
- Brighter: 23–25 mag
- Fainter: 25–27 mag

•**Cross-matched DP0 classifications with "truth" labels** provided in the simulation data.

•**Examined classification accuracy and error rates** as a function of magnitude.

•**Explored improvement strategies** using axis-ratio cuts and machine learning classifiers based on photometric and morphological features.

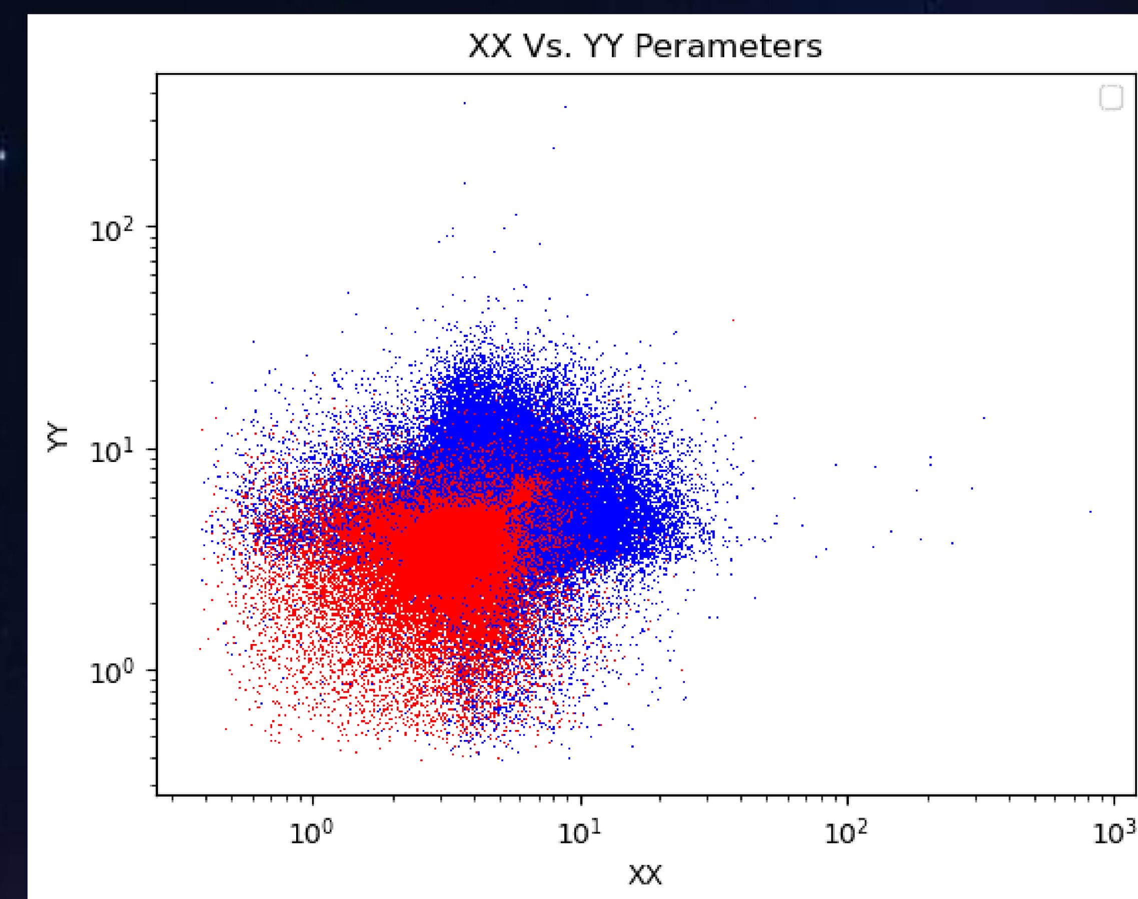
Conclusion

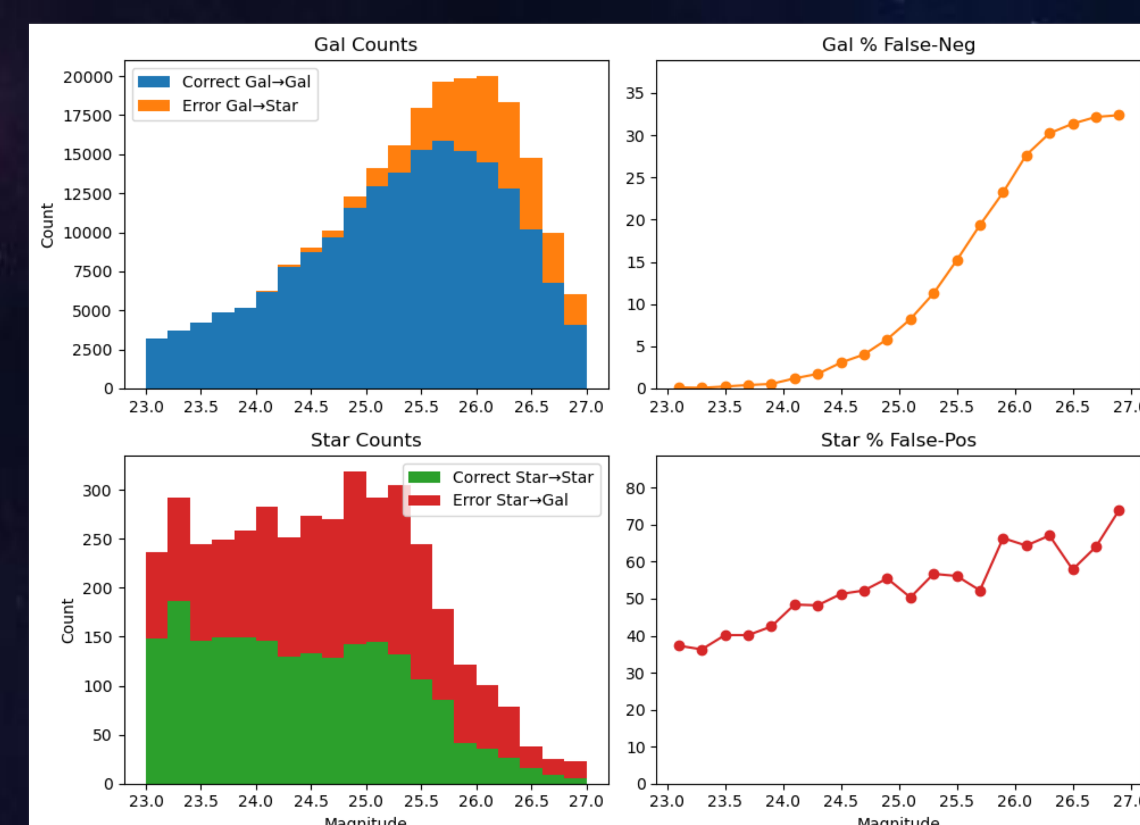
We analyzed 750,000 sources and found star classification accuracy drops below 50% beyond mag 25, while galaxies stay >95% through ~26.5. Shape-based filters recovered 10–15% of misclassified galaxies. A machine learning model using photometry + morphology improved accuracy by 5%.

Future Enhancements:

- Add color features (e.g., g–r, r–i) to better identify faint stars
- Balance training data and tune P(star) thresholds
- Use a hybrid model: HGB for galaxies + secondary classifier for stars

Graphs & Data Analysis





Key Results

Shape-Based Separation (XX vs. YY)

Color Key:

•**Blue**: Real stars (correctly classified)

•**Red**: False stars (misclassified by HGB)

Findings:

Real stars cluster at higher XX (~5–20), YY (~3–15).

False stars (misclassified by HGB) cluster near origin (XX < 5, YY < 3).

Conclusion:

Shape parameters (XX, YY) are strong morphological discriminators—essential for improving classification of faint sources.

discriminator between stars and galaxies.

Magnitude 23–25:

- Stars: Accuracy drops from 60% → 20%
- Galaxies: 100% → 95%

Magnitude 25–27:

- Stars: <40% correct
- Galaxies: 95% → 70%

Takeaway:

Star classification collapses at faint mags; robust galaxy recovery persists to r ≈ 26.5. Improvements are critical beyond 25 mag.

•**Magnitude Ranges:**

•Bright (23–25 mag): ≳ 80% correctly classified

•Intermediate (25–27 mag): Accuracy drops to ~65%

•Faint (25–27 mag): Misclassification exceeds 50%

Morphological Findings:

- Misclassification increases with object roundness ($q \approx 1$)(q≈1), derived from second-moment shape parameters (xx, yy, xy)
- Real structural separation is visible in log(XX) vs. log(YY), motivating shape-based features in HGB

Performance Trade-off:

- Galaxy completeness remains >95% to $r \approx 26.5$r≈26.5
- Star purity collapses past $r \approx 25.5$r≈25.5, with faint stars overwhelmingly misclassified as galaxies

Takeaway:

HGB aggressively pushes galaxy recovery to faint limits but does so at the cost of star classification—especially for round, dim sources.