CORONA: Classification of R Coronae Objects with Neural Network Automation

Maria Dubiel, Ylann Bouis, Emily Bell, Melina Handan, Jonathan Zhang; Mentor: Caitlin Begbie

University of California, Berkeley   Undergraduate Lab at Berkeley, Physics & Astronomy Division

Abstract

R Coronae Borealis stars are a rare type of variable star. Given the rarity of RCB stars (only ∼150 confirmed in the Milky Way), automated detection in large photometric surveys is essential in identifying promising candidates. We introduce CORONA, a neural network classification pipeline aimed at identifying RCB-like light curve signatures in the ROME Galactic Survey.

To overcome the scarcity of labeled RCB data (<100 confirmed RCB light curves from Gaia, OGLE-III, and Karambelkar 2024), we generated an artificial training set of over 1000 light curves. This was achieved by injecting survey-specific noise residuals from ROME data into pre-existing RCB curves and scaling magnitude and error values to accentuate known RCB light curve features. Trained on this extended dataset, the neural network exhibits robust classification between RCB and non-RCB datapoints. Applied on the full dataset, CORONA produces a catalog of probable RCB candidates for spectroscopic follow-up. Our approach streamlines the discovery of rare variable stars in modern-day photometric surveys and can be improved with iterative training using newly confirmed RCB candidates.

Background Information: R Coronae Borealis variable stars

- R Coronae Borealis stars are rare, low-mass, hydrogen-deficient supergiant variable stars [1].
- Characterized by irregular, often abrupt fading events that can reach from 1 to 9 magnitudes fainter than their normal magnitude [2].
- So far, only about 150 RCB stars have been discovered in the Milky Way Galaxy, and several of these are located in the Galactic Bulge [3].
- Since there are confirmed RCB stars in the Galactic Bulge, it is highly likely that our model will find strong RCB candidates in the ROME data.

Figure 1. AAVSO Light Curve of R Coronae Borealis from 1980 to 1990 [4]

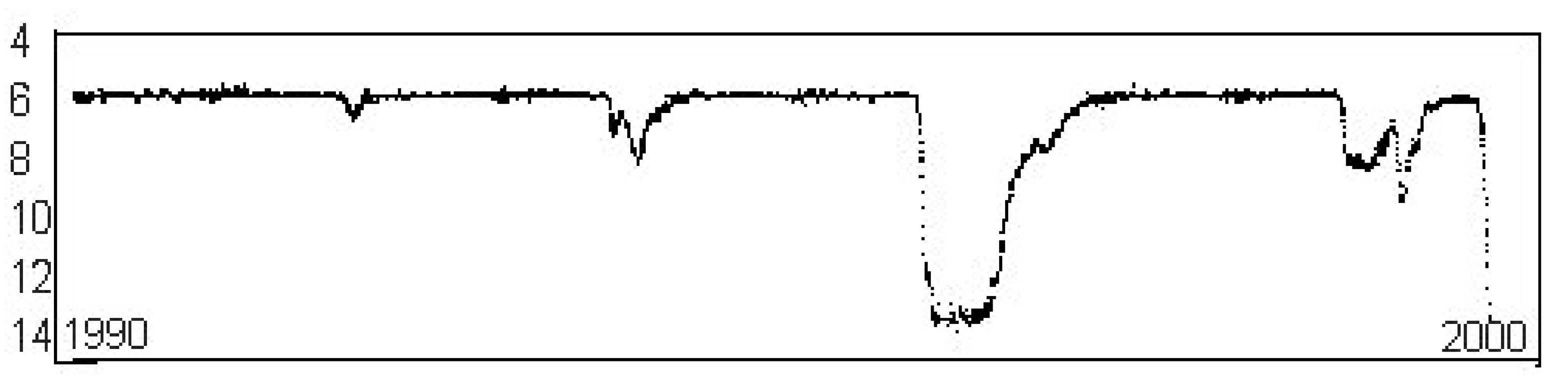

Figure 2. AAVSO Light Curve of R Coronae Borealis from 1990 to 2000 [4]

- Above figures are sample light curves of confirmed RCB. The characteristic dips in magnitude occur at irregular intervals, but they do follow a general shape.
- The light curve will remain flat for several days after a dip as the star remains at that faded magnitude, then start to brighten.

Methodology: Light Curve Manipulation

Process for getting RCB light curves in the ROME format:

- Used data from Gaia and Karambelkar(2024) [5], [6]
- Filtered out variable stars from ROME data and pulled main sequence light curves
- Created a function to insert the noise from the ROME light curves into RCB light curve data
- Five versions of each unique light curve were created, with residuals from different stars

We were only able to obtain fewer than 100 light curves from Gaia, OGLE, and Karambelkar (2024) [5], [7], [6]. So we needed to manipulate our data to obtain more. Reaching 1000 light curves:

- Created function to scale magnitude and magnitude error values
- Created copies scaled by factors of 1.05, 0.95, 1.07, 1.01, 0.88, and 0.85.

Background Information: The ROME Sky Survey

- ROME Sky Survey was conducted by the Las Cumbres Observatory (located in Goleta, California) over a period of three years, from March 2017 to March 2020.
- The purpose of the survey was to look for microlensing events. The observatory surveyed around eight million stars in the Galactic Bulge, and compiled the time series photometry into light curves that show the stars' magnitude over their viewing period [8].

Methodology: The Machine Learning Model

- created a machine learning model that utilized Long Short-Term Memory Networks.
- LSTMs are geared towards time-series data, making them the most suitable to process light curves.
- Used PyTorch.
- The model is designed to be applied to the ROME dataset.
- model is primarily focused on light curve characteristics
- Possible candidates are ranked on the posterior probability of being RCB stars
- Candidates above a certain threshold are compiled and manual analysis methods are then applied, aiming at the verification of our results.

Data

- Collected RCB light curves from: Gaia Alerts [5], OGLE-III Catalog of Variable Stars [7], and Karambelkar (2024) [6]
- Datasets reflect diverse RCB behaviors across instruments and surveys
- Used to train and test the model for generalization
- Enabled model to search the ROME dataset for new RCB candidates

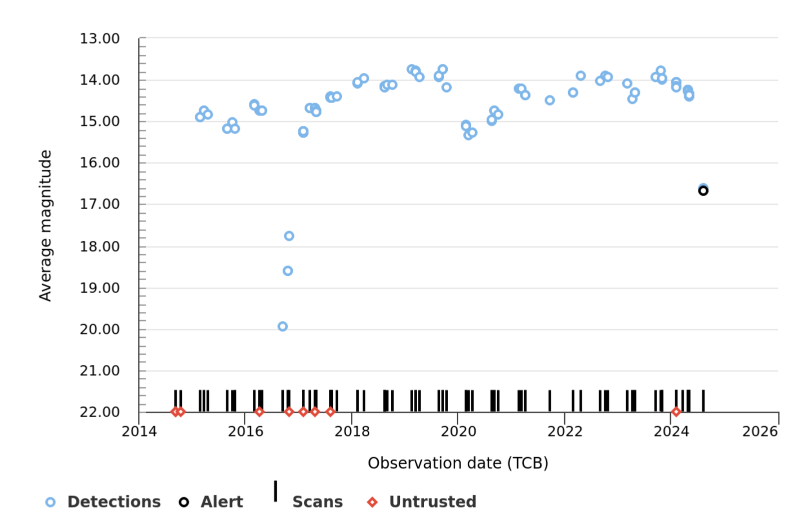

Figure 3. RCB light curve collected from Gaia [5]

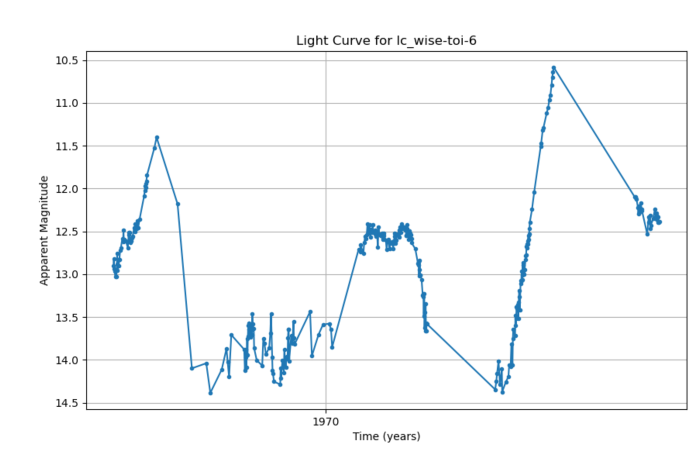

Figure 4. RCB light curve from "An infrared search for R Coronae Borealis Stars" [9]

Future Work

- Running ROME data through the model to identify stars with high probabilities of being RCB stars and adding those back into the training data for further improvement.
- Fix errors in the DELightCurveSimulation package and use the corrected package to create a larger and more accurate set of training data.
- Use time on NERSC to run the entire set of ROME light curves through our model. The stars with the highest probability of being an RCB will then be checked by hand.
- Make model open source, which will hopefully assist future researchers.

Conclusion

- Trained on real and simulated light curves
- 42000 main sequence curves (42 unique)
- 2468 RCB curves (all unique, real and simulated)

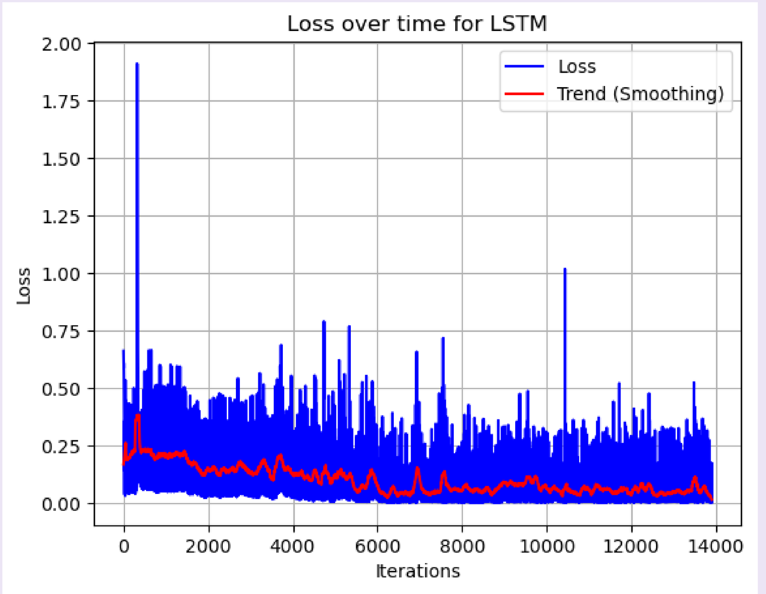

Figure 5. Loss over time for LSTM model

	RCB Star	Main Sequence Star
Probability (avg.)	0.997	0.0246

Table 1. Predicted probabilities from our model for each kind of star.

- When given an RCB star, it outputs an average probability of 99%.
- When given a Main Sequence star, it outputs an average probability of 0.0246 of being an RCB Star (good).
- Both of these figures may seem good, but they suggest room for improvement (reducing overfit and more data).
- Overall with more training, the model seems pretty good on these datasets.

Acknowledgments

We would like to thank the Undergraduate Laboratory at Berkeley (ULAB) for fostering this project and providing a supportive environment for undergraduate research. We also gratefully acknowledge Las Cumbres Observatory for providing access to the ROME dataset, which made this work possible.

References

[1] C. at Harvard, *R coronae borealis stars*, Sensing the Dynamic Universe, CFA at Harvard, 2020.

[2] B. E. Schaefer, "Light curves for 10 r coronae borealis stars for longer than a century: Secular evolution, dip statistics, and a general model for the shape of isolated light-curve dips," *Monthly Notices of the Royal Astronomical Society*, vol. 527, no. 3, 2024.

[3] P. Tisserand, J. Marquette, P. R. Wood, *et al.*, "R coronae borealis stars in the galactic bulge discovered by eros-2," *Astronomy & Astrophysics*, vol. 481, no. 3, pp. 673–690, 2008.

[4] K. Davis, "R coronae borealis," *The American Association of Variable Star Observers*, 2000.

[5] U. Institute of Astronomy University of Cambridge, *Gaia alerts index*, 2015.

[6] A. Karambelkar *et al.*, "An infrared census of r coronae borealis stars ii—spectroscopic classifications and implications for the rate of low-mass white dwarf mergers," *Publications of the Astronomical Society of the Pacific*, vol. 136, no. 8, 2024.

[7] I. Soszynski *et al.*, "The optical gravitational lensing experiment. the ogle-iii catalog of variable stars. r coronae borealis stars in the large magellanic cloud," *Acta Astronomica*, 2009.

[8] R. A. Street, E. Bachelet, Y. Tsapras, *et al.*, "Rome/rea: Three-year, tri-color timeseries photometry of the galactic bulge," *Publications of the Astronomical Society of the Pacific*, vol. 136(6), 064501, 2024.

[9] A. Karambelkar *et al.*, "Census of r coronae borealis stars. i. infrared light curves from palomar gattini ir," *The Astrophysical Journal*, vol. 910, no. 10, 2021.

[10] A. Karambelkar, "An infrared search for r coronae borealis stars," *Zenodo*, 2024.