Classifying the Life Stage of Stellar Objects: from Birth to Death

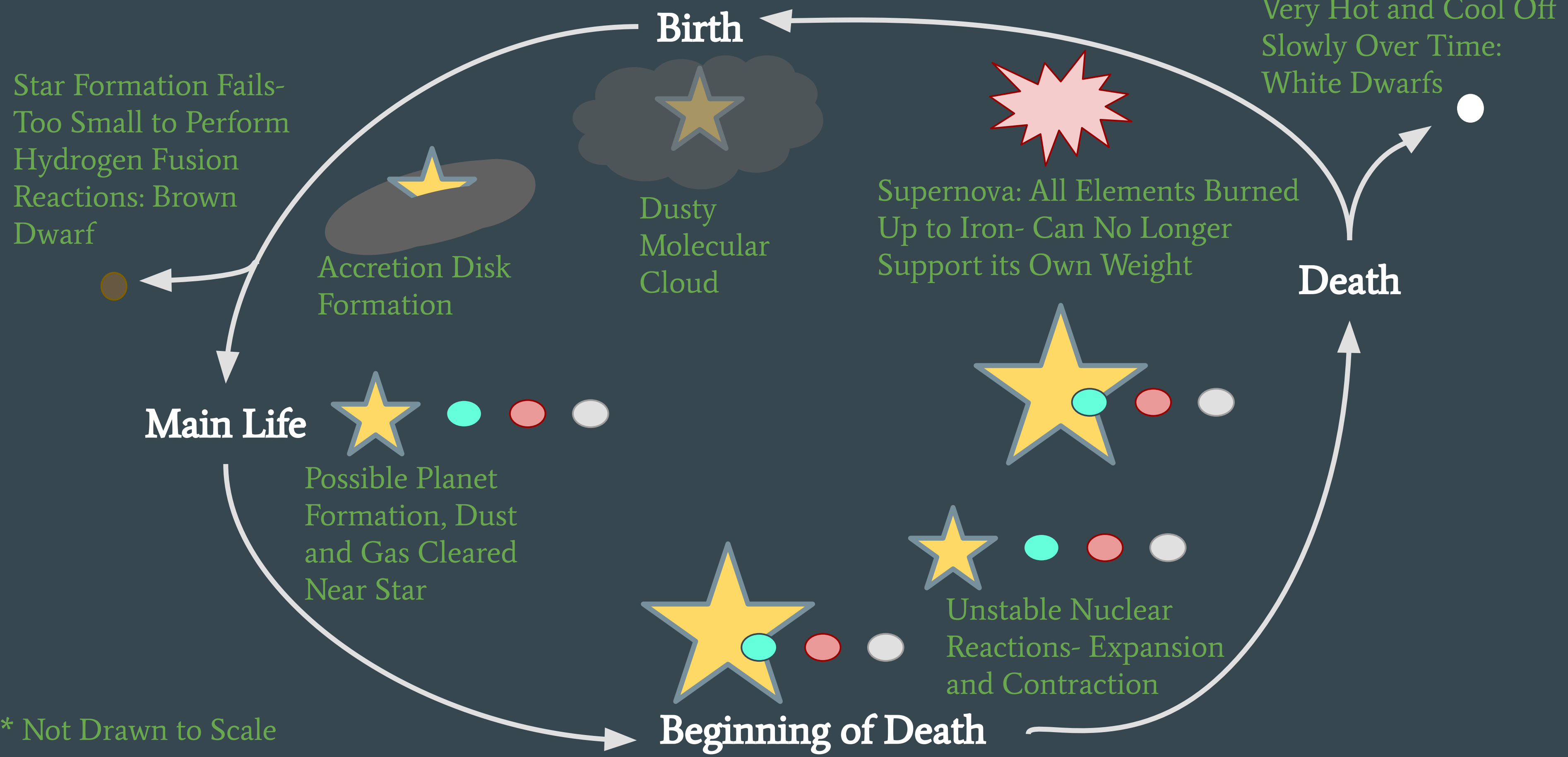
By: Hermione(Zhiyan) Deng, Elizabeth Hora, Vanessa Venkataraman

Agenda

- Background on Astronomy Concepts / Explanation of Data
- Exploratory Data Analysis
- Classification Models
- Future Applications of Our Work

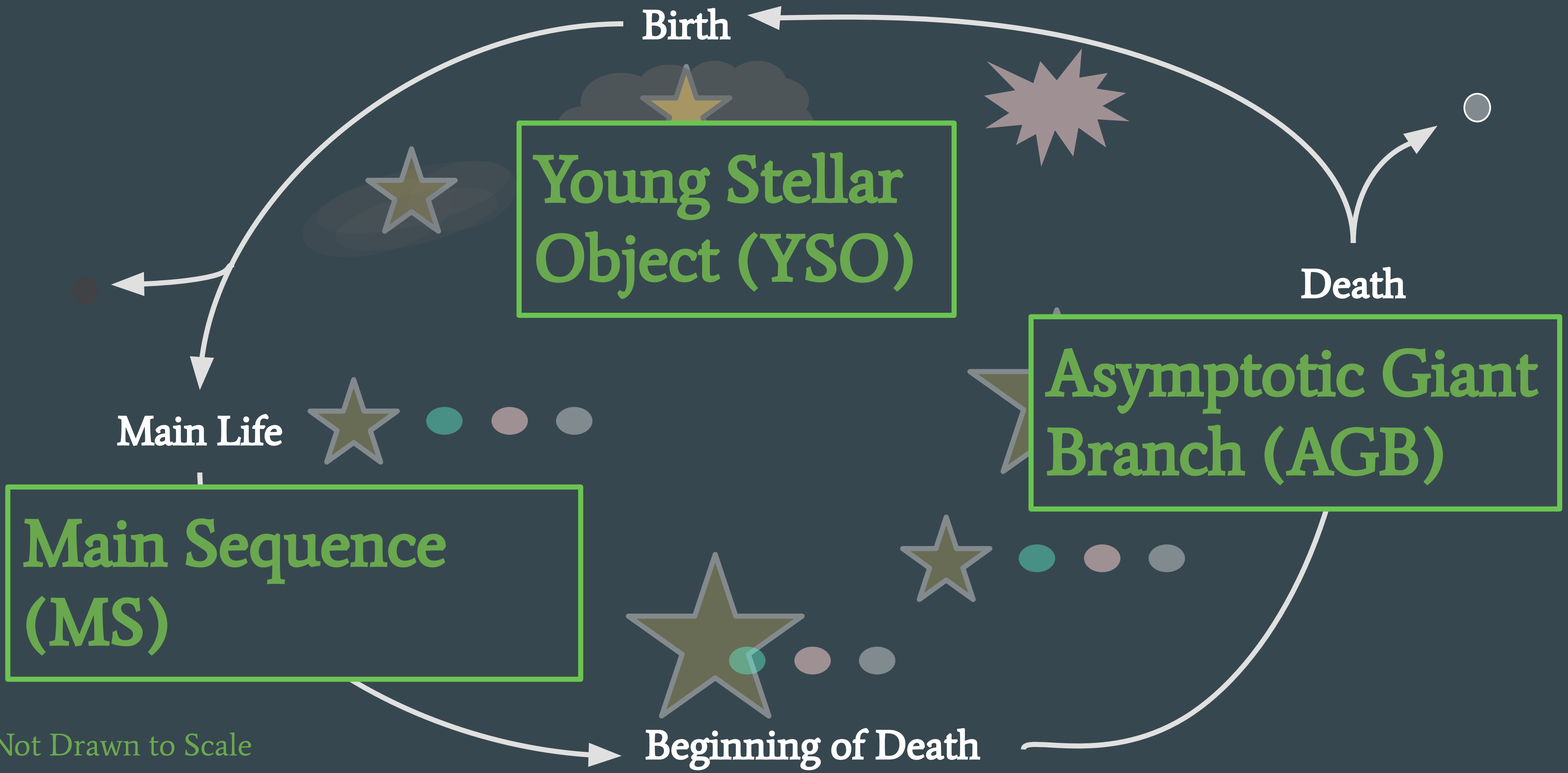
Background on Astronomy Concepts / Explanation of Data

Life Cycle of a Star



* Not Drawn to Scale

Our Objective: 3 Main Classifications (Based on Hertzsprung-Russell (HR) Diagram)



* Not Drawn to Scale

Types of Measurements- Part of the SPHEREx Mission Catalog

Gaia Telescope:
0.4, 0.6, 0.8
Microns
(Space-Based)

2MASS Telescope:
1.2, 1.6, 2.2
Microns
(Ground-Based)

WISE Telescope:
3.4, 4.6, 12, 22
Microns
(Space-Based)

Visible Light

Infrared (IR) Spectrum

Why Use Photometry if Spectroscopy Provides the Right Answer?

Time Constraints for Each Method

Number of Objects in the WISE Database: 750 million

Spectroscopy:

Number of Minutes of Observing per Object: ~30 minutes

Amount of Time to Observe Every Object: **~ 43,000 years**

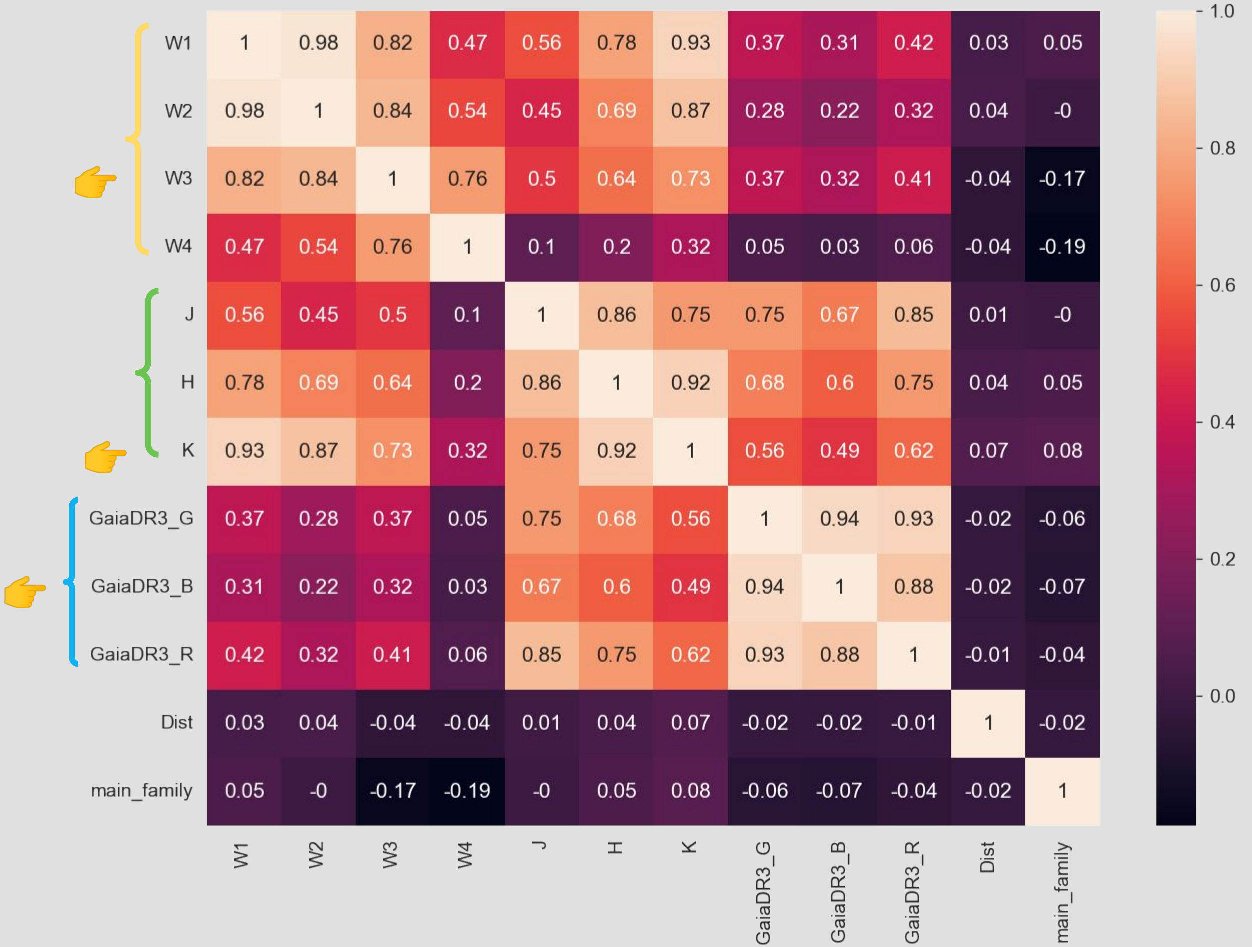
Photometry:

Amount of Time to Classify Every Object: **~ 0.5 years** (Space-Based; WISE)

or **~ 2-3 years** (Ground-Based; 2MASS)

EDA

Variables


W1 1 0.98 0.82 0.47 0.56 0.78 0.93 0.37 0.31 0.42 0.03 0.05
W2 0.98 1 0.84 0.54 0.45 0.69 0.87 0.28 0.22 0.32 0.04 -0
W3 0.82 0.84 1 0.76 0.5 0.64 0.73 0.37 0.32 0.41 -0.04 -0.17
W4 0.47 0.54 0.76 1 0.1 0.2 0.32 0.05 0.03 0.06 -0.04 -0.19
J 0.56 0.45 0.5 0.1 1 0.86 0.75 0.75 0.67 0.85 0.01 -0
H 0.78 0.69 0.64 0.2 0.86 1 0.92 0.68 0.6 0.75 0.04 0.05
K 0.93 0.87 0.73 0.32 0.75 0.92 1 0.56 0.49 0.62 0.07 0.08
GaiaDR3_G 0.37 0.28 0.37 0.05 0.75 0.68 0.56 1 0.94 0.93 -0.02 -0.06
GaiaDR3_B 0.31 0.22 0.32 0.03 0.67 0.6 0.49 0.94 1 0.88 -0.02 -0.07
GaiaDR3_R 0.42 0.32 0.41 0.06 0.85 0.75 0.62 0.93 0.88 1 -0.01 -0.04
Dist 0.03 0.04 -0.04 -0.04 0.01 0.04 0.07 -0.02 -0.02 -0.01 1 -0.02
main_family 0.05 -0 -0.17 -0.19 -0 0.05 0.08 -0.06 -0.07 -0.04 -0.02 1
W1 W2 W3 W4 J H K GaiaDR3_G GaiaDR3_B GaiaDR3_R Dist main_family
1.0
0.8
0.6
0.4
0.2
0.0

Data Processing- Faint Objects

- Missing measurement data are due to objects being too faint to observed, not because they are not there

~~.dropna()~~ ➡️ fillna()

Data Processing- Eliminating Uncertain Distances

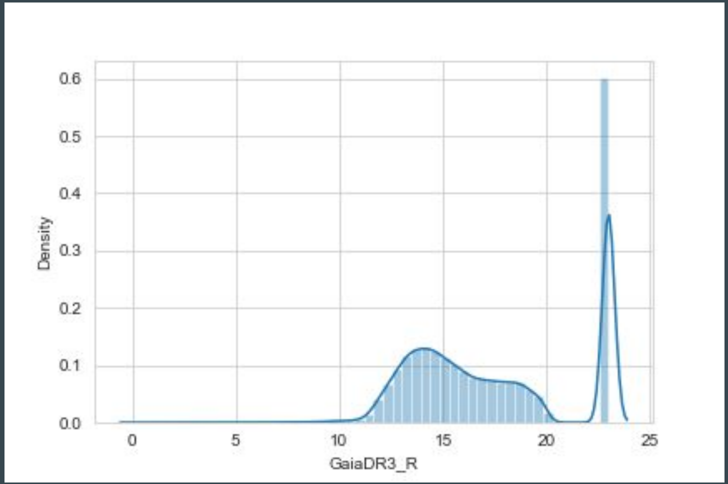
What's the Problem?

Not every object can have its distance calculated reliably.

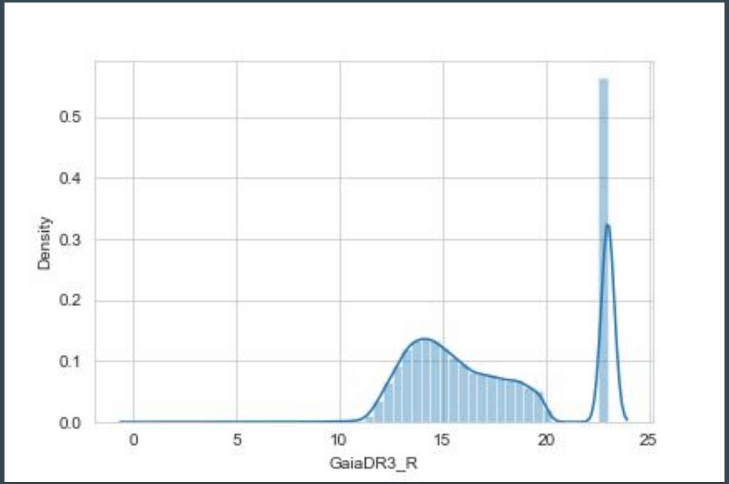
What's the Solution?

We drop distances that have a value of "NA"

All Objects



Objects with a Known Distance



Classification Models

Logistic Regression – with all bands

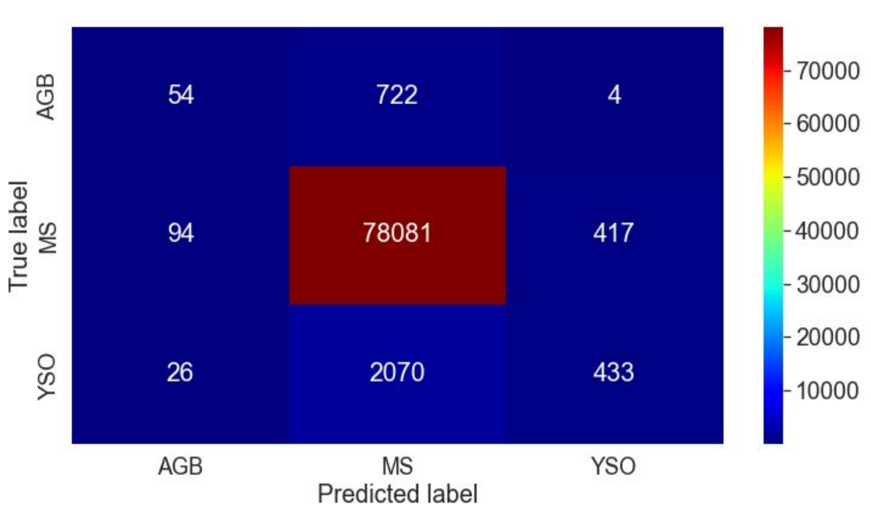
Step 1: The ideal C-Value (hyperparameter): **0.48**

Step 2: The logistic regression

Step 3: The accuracy score: **0.96**

Step 4: The Confusion Matrix

Confusion Matrix with ALL Data

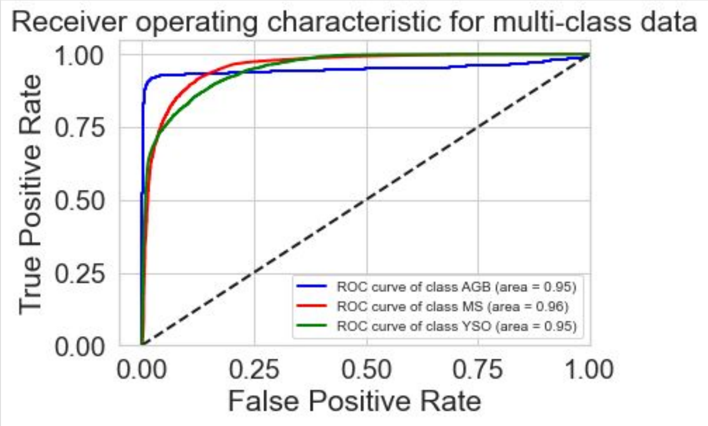


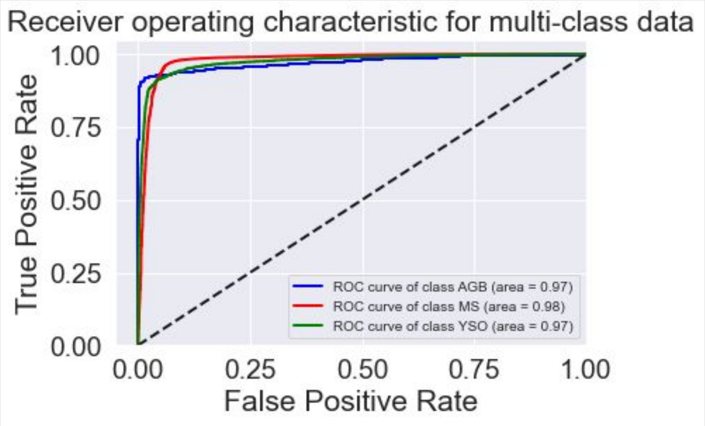
ROC & AUC from **SVM** Model

All

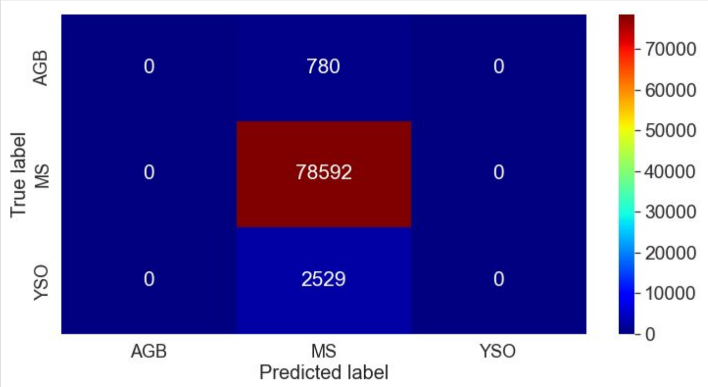
W3, K & GAIA-B

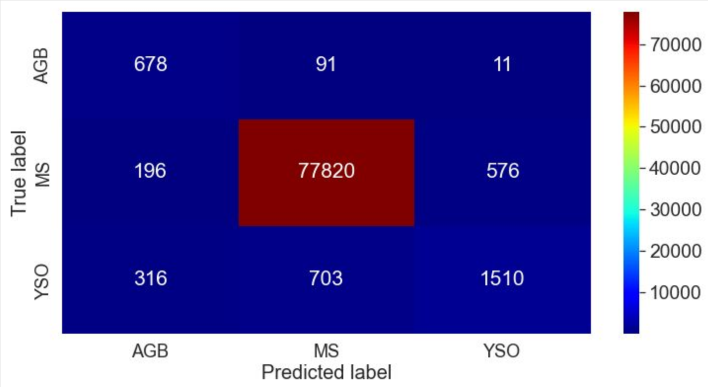
OneVsRestClassifier



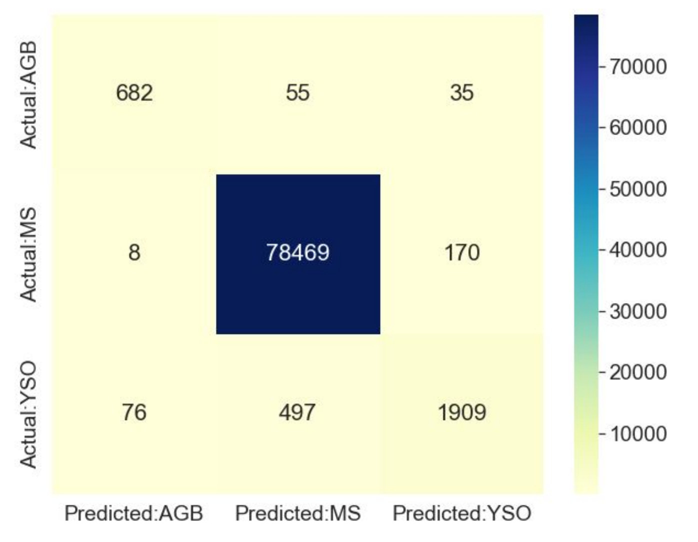


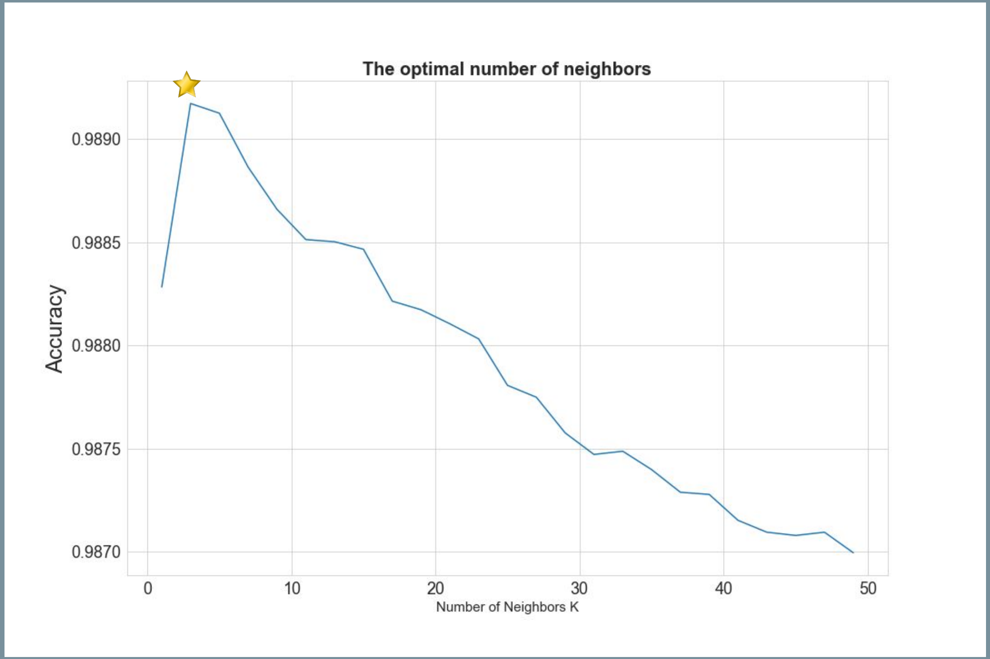
Confusion Matrix





K-Nearest Neighbors





Classification Report:

	precision	recall	f1-score	support
AGB	0.89	0.88	0.89	772
MS	0.99	1.00	1.00	78647
YSO	0.90	0.77	0.83	2482
accuracy			0.99	81901
macro avg	0.93	0.88	0.90	81901
weighted avg	0.99	0.99	0.99	81901

Based on cross-validation, the optimal number of neighbors = 3
The cross-validation at 3 neighbors = ~98.91%

Beyond our Results- Future Work

Future Applications of Our Work

- Identifying AGB and YSO stars
  - We can apply our model on new data to select which stars to take spectra of to confirm their identity
  - Several thousand objects is a more manageable search than millions
- Identify Outliers
  - For the misclassified objects, follow up and see why they are different

Sources of Our Data- Part of the SPHEREx Mission Catalog

Alksnis et al., “General Catalog of Galactic Carbon Stars by C. B. Stephenson. Third Edition.” Baltic Astronomy, 10, 1, 2001, 10.1515/astro-2001-1-202.

Kuhn et al., “The Effect of Molecular Cloud Properties on the Kinematics of Stars Formed in the Trifid Region.” ApJ, 937, 46, 2022, 10.3847/1538-4357/ac6fe8.

Li et al., “Carbon Stars Identified from LAMOST DR4 Using Machine Learning.” ApJS, 234, 31, 2018, 10.3847/1538-4365/aaa415.

Huchra et al., “THE 2MASS REDSHIFT SURVEY—DESCRIPTION AND DATA RELEASE.” ApJS, 199, 26, 2012, 10.1088/0067-0049/199/2/26.

Ishihara et al. “Galactic distributions of carbon- and oxygen-rich AGB stars revealed by the AKARI mid infrared all-sky survey.” A&A, 534, A79, 2011, doi.org/10.1051/0004-6361/201117626.

Prusti et al., “The Gaia Mission.” A&A, 595. 2016, doi.org/10.1051/0004-6361/201629272.

Skrutskie et al., “The Two Micron All Sky Survey (2MASS).” AJ, 131, 1163, 2006, 10.1086/498708.

Suh, K.-W. “A New Catalog of Asymptotic Giant Branch Stars in Our Galaxy.” ApJS, 256, 43, 2021, 10.3847/1538-4365/ac1274.

Wright et al., THE WIDE-FIELD INFRARED SURVEY EXPLORER (WISE): MISSION DESCRIPTION AND INITIAL ON-ORBIT PERFORMANCE.” AJ, 140, 1868, 2010, 10.1088/0004-6256/140/6/1868.

Additional Sources

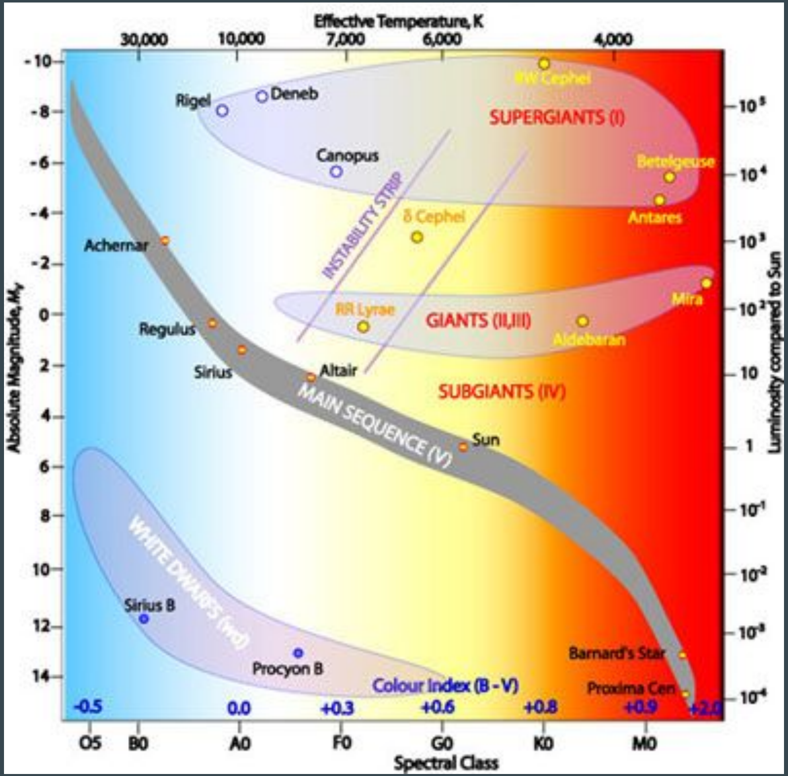
Fujii et al., “BVRIJHK photometry of post-AGB candidates.” A&A, 385, 2002, 10.1051/0004-6361:20020178.

Groenewegen and Blommaert. “Near-infrared photometry and optical spectroscopy of IRAS sources in the Small Magellanic Cloud.” A&A, 332, 1998.

Testor et al., “Optical and infrared observations of the young SMC blob N26 and its environment.” A&A, 564, A31, 2014, https://doi.org/10.1051/0004-6361/201118484.

Appendix

Hertzsprung-Russell Diagram



- This diagram relates the absolute magnitude (with respect to Vega) with the relative brightness compared to the Sun
- The spectral class has to do with which color the object appears, which is strongly related with temperature
- Meaning of Asymptotic Giant Branch (AGB stars)

Sources of Our Data- Part of the SPHEREx Mission Catalog

Star Classification	Method	Description
YSO	Photometry	Examined Deep Molecular Clouds (clusters of star forming regions), compared the SEDs to YSO models
MS	Photometry/Spectroscopy	Known MS stars of every star type (O, B, A, F, G, K, M) were queried from Simbad (Astronomy Database)
AGB	Spectroscopy	Looked at AGB candidates and analyzed obtained spectra

What are the Differences between Spectroscopy and Photometry? What are Some Challenges Associated with Each Measurement Type?

Main Challenges- Spectroscopy versus Photometry

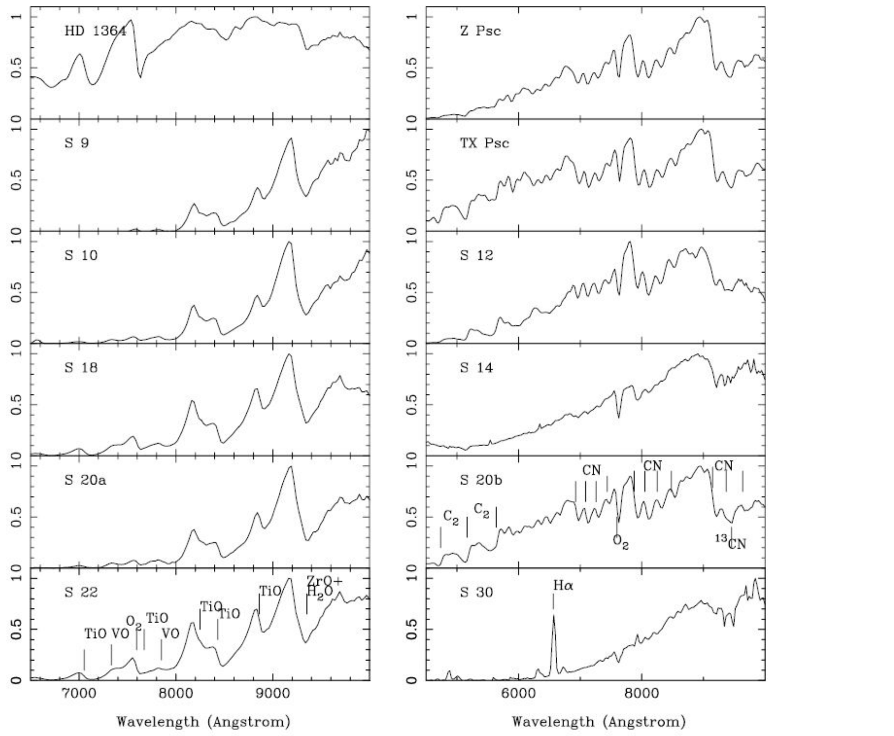
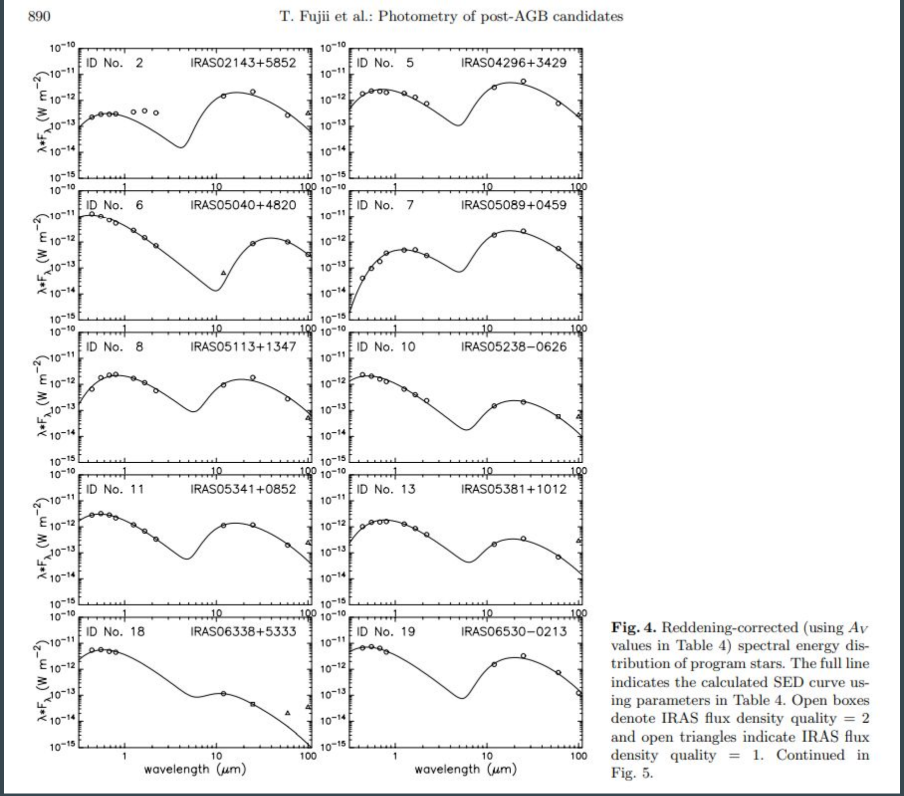


Fig. 1. The optical spectra of the spectral standards and some SMC objects. To the left the oxygen-rich objects, to the right the carbon-rich objects. Some prominent lines are indicated, including telluric O_2 and H_2O.

- Example of Oxygen-Rich versus Carbon-Rich AGB stars
- Emission lines determine the temperatures and composition of objects
- Observations taken from the ESO/MPI 2.2m telescope at La Silla, Chile (ground-based)

Main Challenges- Spectroscopy versus Photometry



- Locating Candidate AGB Stars
- Points are Measurements of Bands: B, V, R, I, J, H, K, IRAS bands (12, 25, 60, 100 microns)
  - IRAS (space-based telescope) predates and does not have as good as a resolution as WISE
- Lines are the Spectral Energy Distribution (SED) for Each Object (modeled against a grid, takes seconds to find an appropriate fit per object)
- Observations taken from the Kiso Observatory (ground-based), University of Tokyo

What are Some Challenges with Photometry?

Main Challenges- Minor Differences in SEDs

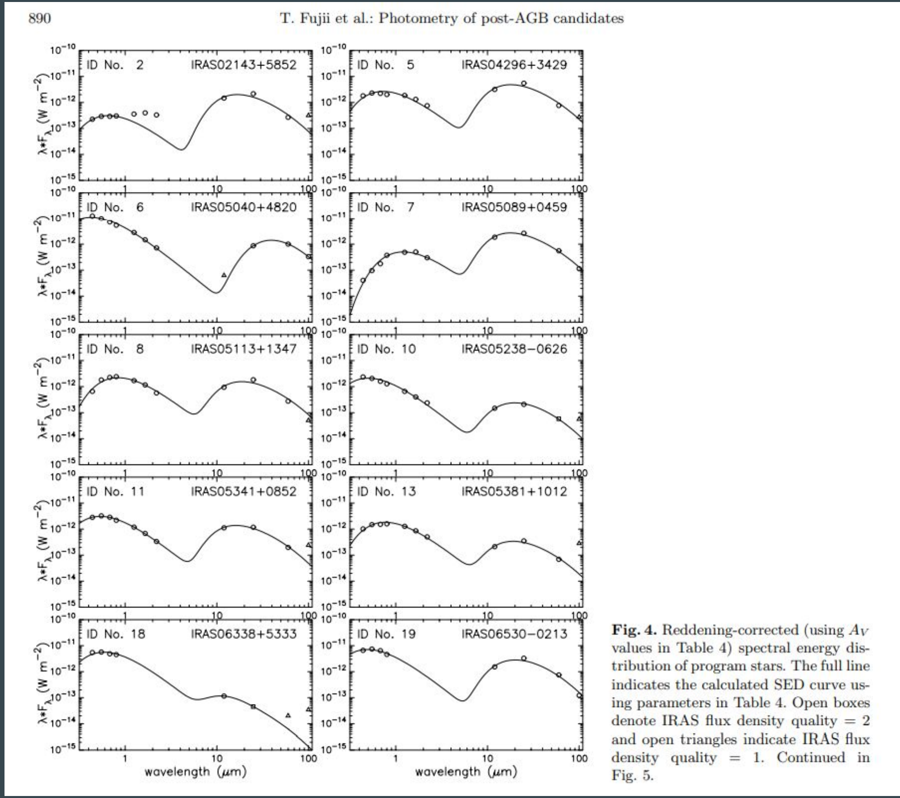
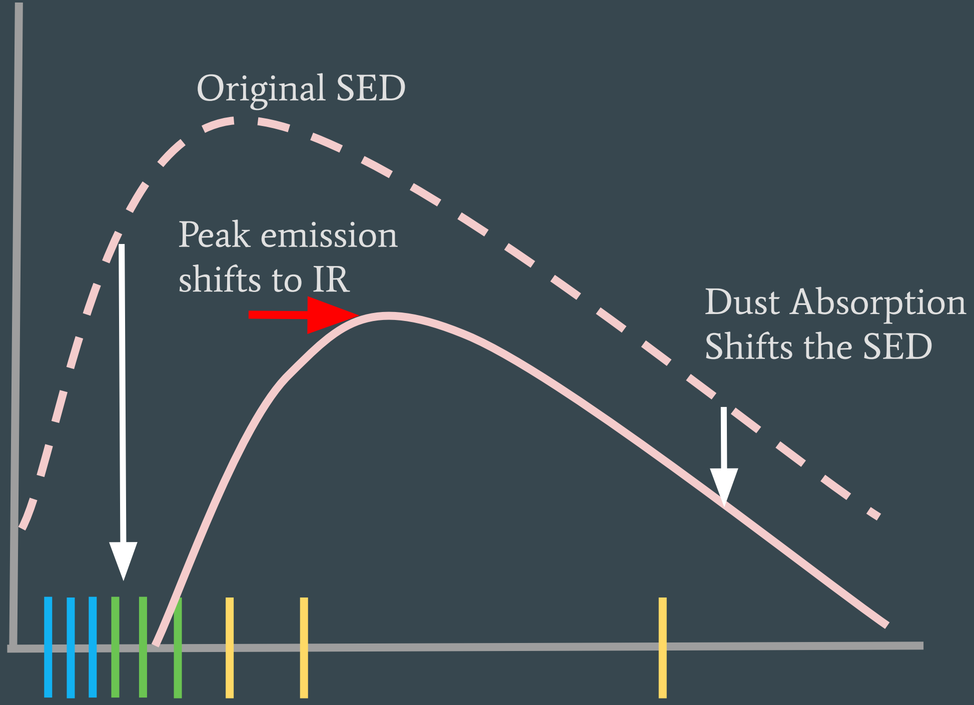


Fig. 4. Reddening-corrected (using A_V values in Table 4) spectral energy distribution of program stars. The full line indicates the calculated SED curve using parameters in Table 4. Open boxes denote IRAS flux density quality = 2 and open triangles indicate IRAS flux density quality = 1. Continued in Fig. 5.

- Many of the curves look similar
- The SED is no substitute for a spectra

Extinction Observed in Measurements



- Extinction describes the light absorbed by dust
  - Dust affects optical more than infrared, which can make objects appear redder
- With a limited number of bands, this could be challenging to overcome
- Analogous situation: firefighters looking for people in a smoke-filled room using their eyes versus an infrared camera

* Example Based on (Testor et al., 2014)