# Discovering Ultracool Dwarfs in Deep Surveys using Machine Learning Methods

Malina Desai

UC San Diego
CASS

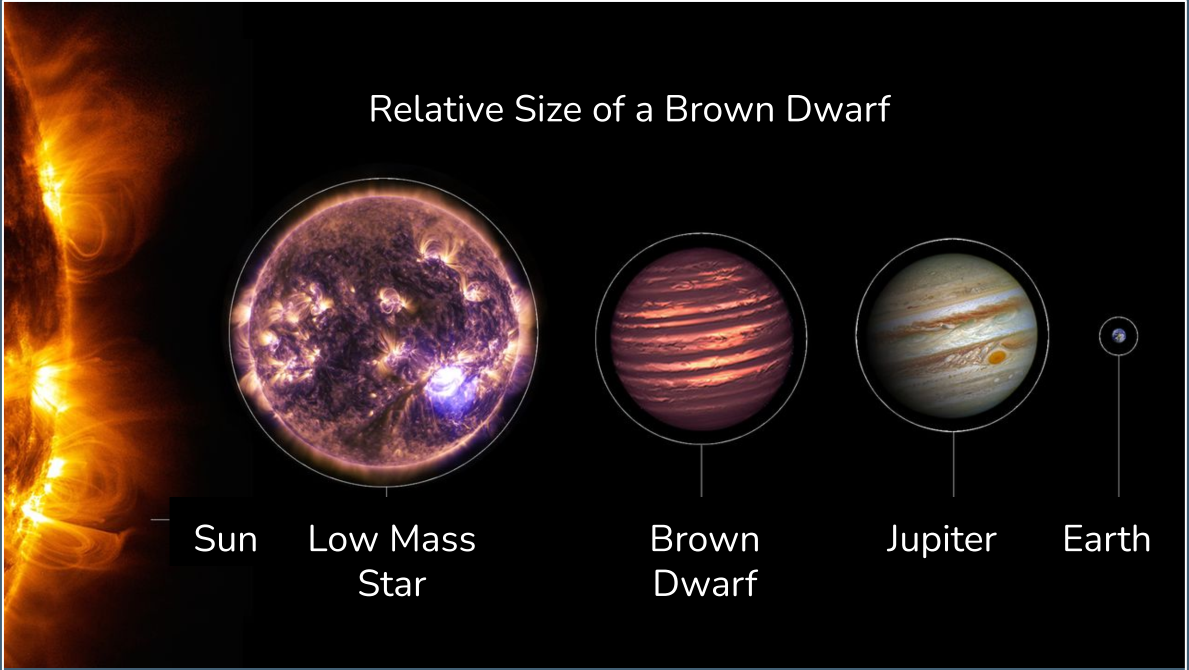
# Overview

UC San Diego

CASS

# What are Ultracool Dwarfs?

UC San Diego
CASS




Artist rendition of relative size of a brown dwarf to the Sun, a low-mass star, Jupiter, and Earth with sizes to scale.

- Spectral class M6 and later (M, L, T, Y)
- Late L, T, & Y form below the hydrogen burning mass limit (brown dwarfs)
- Radiate in the infrared
- Common objects

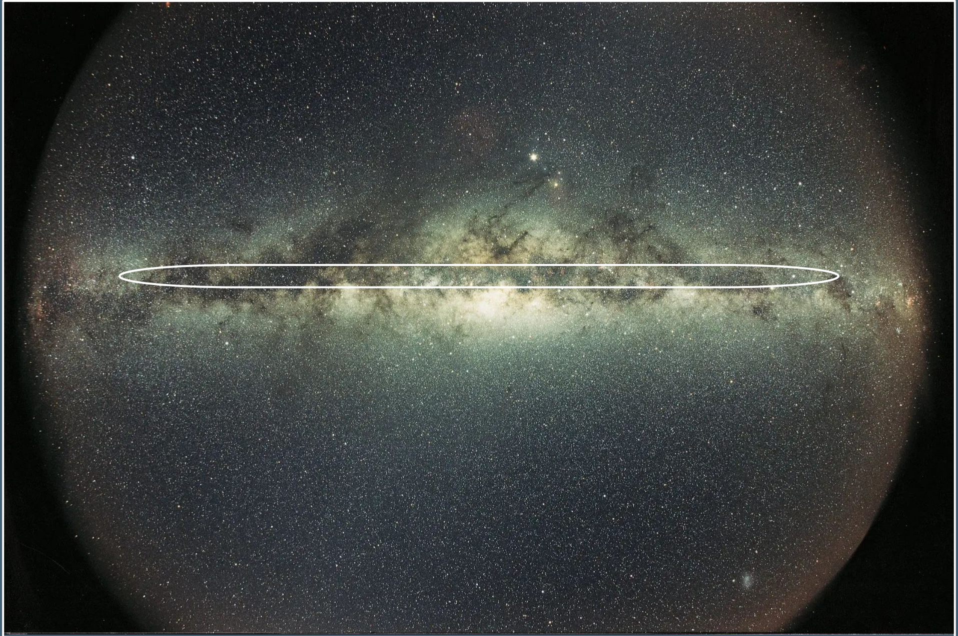
$T_{eff} \leq 3000°K$
$Mass \leq 0.1\ M_\odot$

Image Credit: Jupiter: NASA, ESA, A. Simon (NASA, GSFC); Sun and Low-Mass Star: NASA, SDO; Brown Dwarf: NASA, ESA, JPL-Caltech; Earth: NASA; Infographic: NASA, E. Wheatley (STScI).

# Why do we want to find them?

UC San Diego
CASS

- Understanding differences between star and planet formation
- Improve current stellar models
- Increase known sources at further distances (> 100 parsecs)
- Improve measurement of Milky Way parameters such as thin disk scaleheight

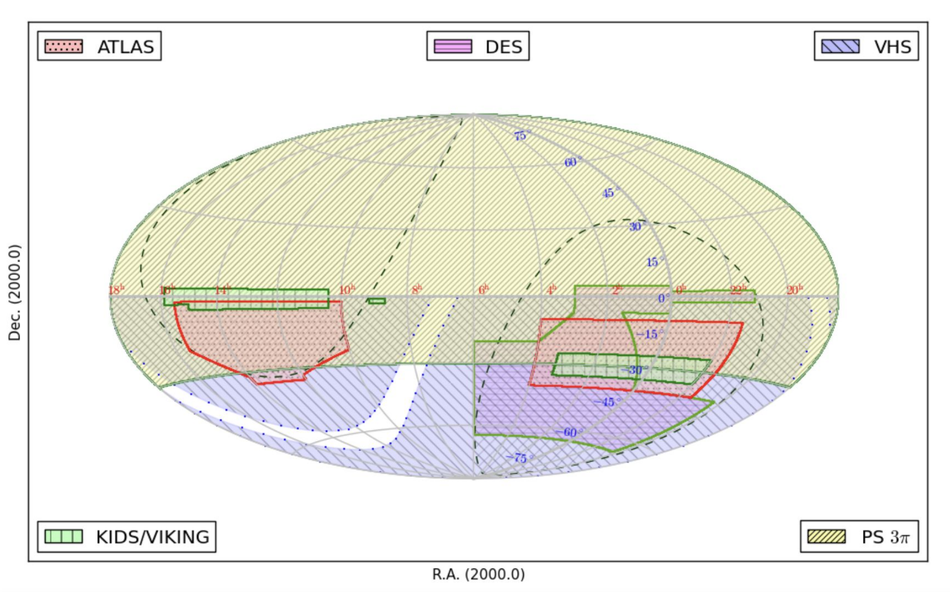


Milky Way with thin disk highlighted

Image Credit: © *Dirk Hoppe*

# Wide-Field Surveys

- PanSTARRS, Sloan Digital Sky Survey (SDSS), and ATLAS in the optical
- 2-Micron All Sky Survey (2MASS) in the near-infrared
- Wide-field Infrared Survey Explorer (WISE) catalogue




Coverage of Pan-STARRS in yellow, DES in purple, VHS in blue (Shanks 2016)

# Deep Sky Surveys

## Dark Energy Survey (DES)

- Search for galaxies and aims to constrain dark energy
- Depth: 21.7 in mags in near-infrared Y band
- 5,000 $deg^2$ coverage in South

## VISTA Hemisphere Survey (VHS)

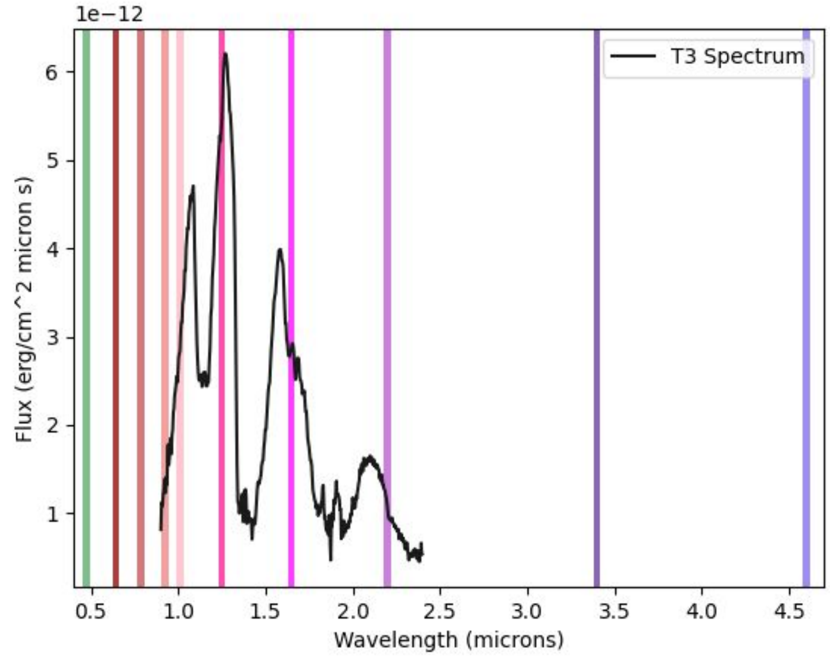
- Search for high-redshift quasars, low mass-stars
- 30 x deeper than 2MASS
- 21,000 $deg^2$ coverage in South

## CatWISE2020 Catalogue

- Known for discovering brown dwarfs
- Depth: W1 = 17.7 mag and W2 = 17.5 mag

# Photometric Bands

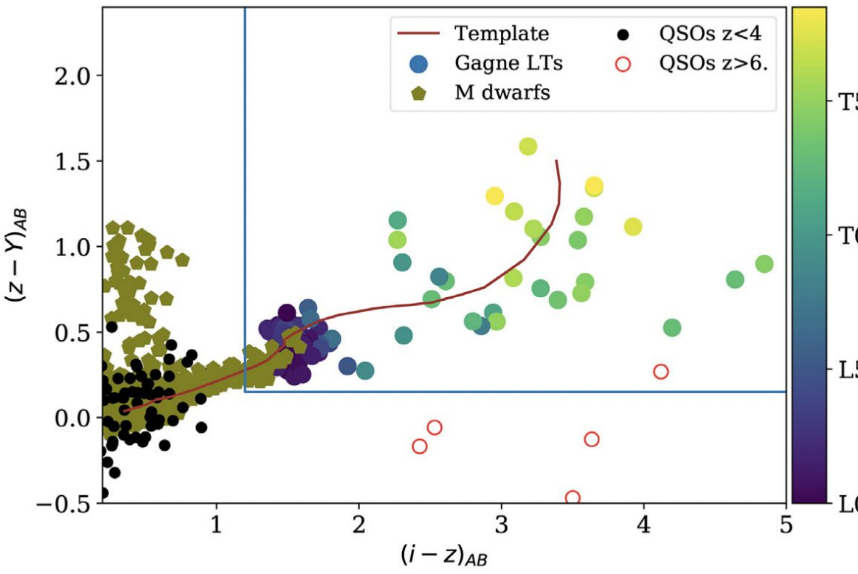
- PanSTARRS and DES survey use grizY bands
- VHS has many sources in J and $K_s$ bands, analogous to 2MASS J and $K_s$
- CatWISE has W1, W2 bands




In order of highlights: g, r, i, z, Y, J, H, $K_s$, W1, and W2 bands overlaid on a T3 Spectrum

# Previous Research

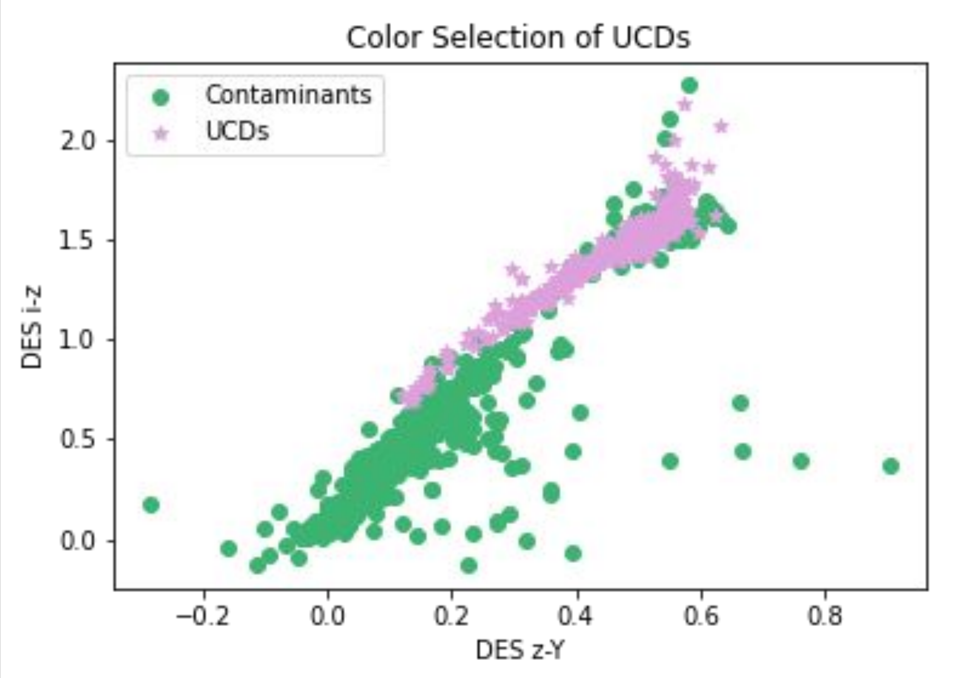
- Carnero Rosell et al. 2019 created a catalogue of 11,745 brown dwarfs of type L0 to T9 using DES, VHS, and WISE photometry
- Used photometric color selection to identify possible candidates
- Spectral classification based on color templates
- Measured thin disk scale height to be 450 pc




Color cut displaying UCDs and quasars

# Color Selection

- Color: magnitudes in two photometric bands subtracted from each other
- The color-color plot shows some overlap between UCDs and our contaminant sample
- ML can help find patterns of all color combinations to identify UCDs




z-Y color plotted against i-z color for UCDs and contaminants

# Limitations

01
## Color cut

Have to pre-determine expected color

02
## Large data

Time–consuming to test several cuts on large datasets

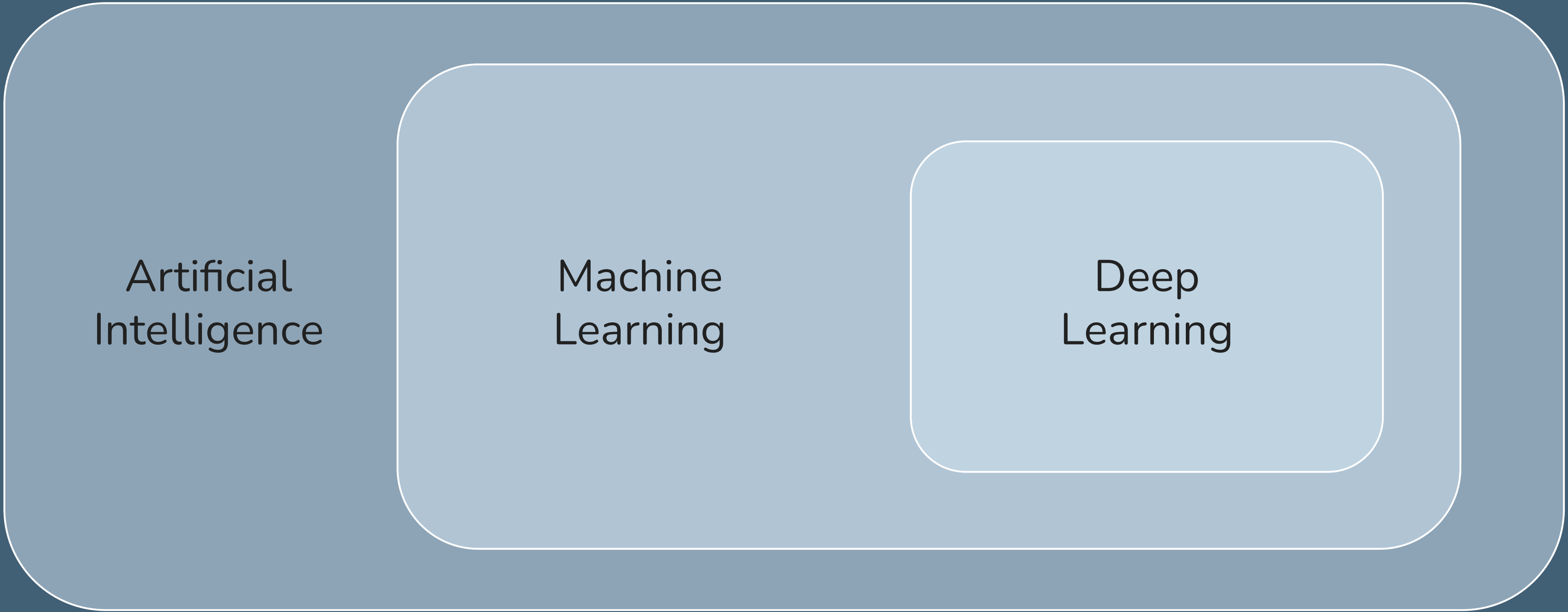
03
## Limited photometry

Hard to incorporate many bands and quickly add new photometry

04
## Contaminants

Can improve on methods used to extract contaminants

# What is Machine Learning?



# What is Machine Learning?

Artificial Intelligence

Machine Learning

- Set of algorithms and statistical models that identify patterns in data
- Includes Random Forests

# What are Random Forests (RF)?

A set of decision trees that is built on “features” and outputs “labels”

| | |
|---|---|
| Training | Subset of pre-classified data used to identify most accurate decision trees |
| Test | separate subset of pre-classified data used to determine tree performance |

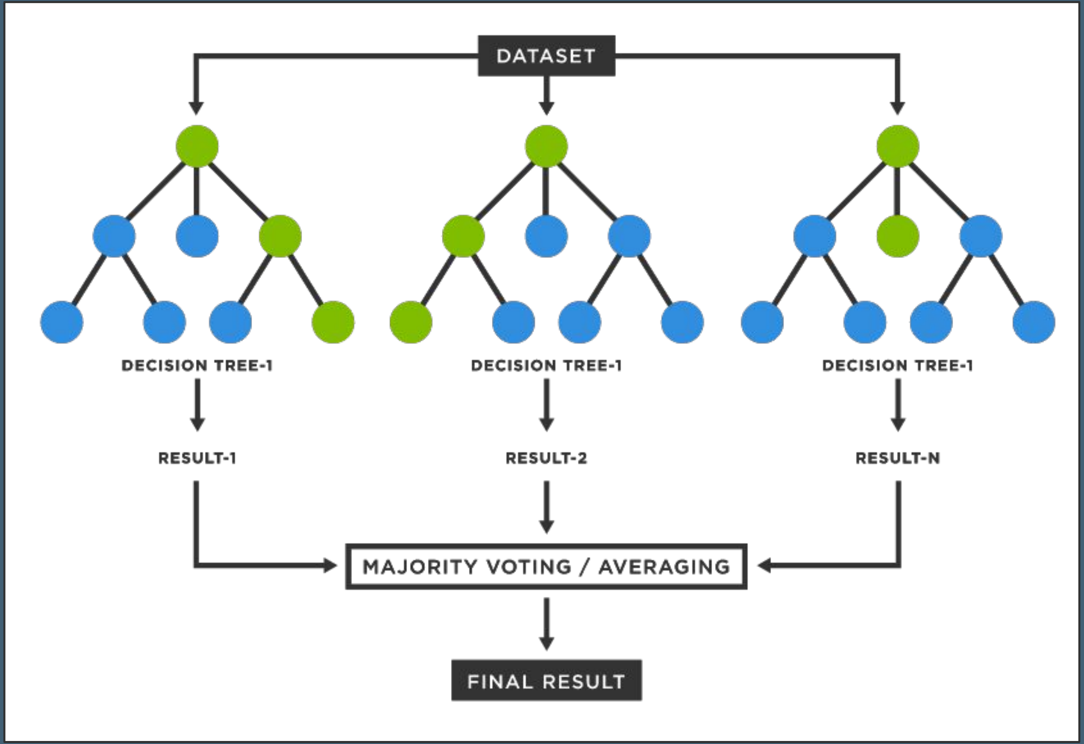



Image Credit: https://www.tibco.com/reference-center/what-is-a-random-forest

# Inputs and Outputs

| $f_1$ | $f_2$ | $f_3$ | Label |
|---|---|---|---|
| 1 | 3 | 5 | 0 |
| 2 | 4 | 6 | 1 |
| 5 | 6 | 9 | 0 |
| 7 | 8 | 0 | 1 |
| 4 | 6 | 8 | ? |

Training

Test

Features

| Prediction |
|---|
| 0 |
| 1 |
| 0 |
| 0 |
| 1 |

# Measuring RF Performance

| | |
|---|---|
| True Positives (TP) | False Positives (FP) |
| False Negatives (FN) | True Negatives (TN) |

| | | |
|---|---|---|
| **Precision** | Measure of misclassified objects of one class | $\frac{TP}{TP + FP}$ |
| **F1 Score** | Balanced measure of model performance | $\frac{TP}{TP + \frac{1}{2}(FP + FN)}$ |

# Why RFs?

01

## Multiple Data Points

Can investigate all colors to identify objects and rank their importance

02

## Efficiency

ML methods can sort through millions of sources at a time

03

## Contaminants

Filter galactic and high-redshift objects

04

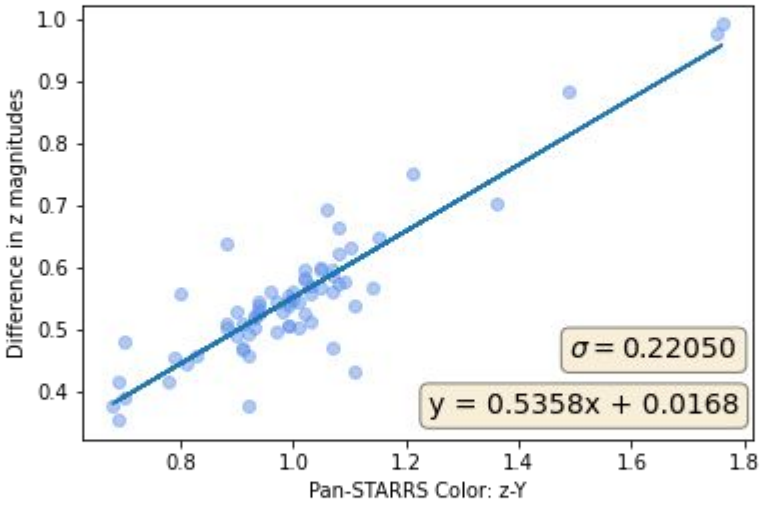
## Previous Research

RFs can identify brown dwarf candidates based on photometric colors and indices (Aganze et al. 2022, Gong et al. 2022)

# Photometric Correction

| | | |
|---|---|---|
| PanSTARRS | ⇨ | DES |
| 2MASS | ⇨ | VHS |
| AllWISE | ⇨ | CatWISE |

# Photometric Correction




Correction of DES z-band magnitudes

- All photometry in the Ultracool Sheet is photometrically corrected to deep sky surveys using linear regression based on color
- MSE of fits: 0.0083 - 0.0019

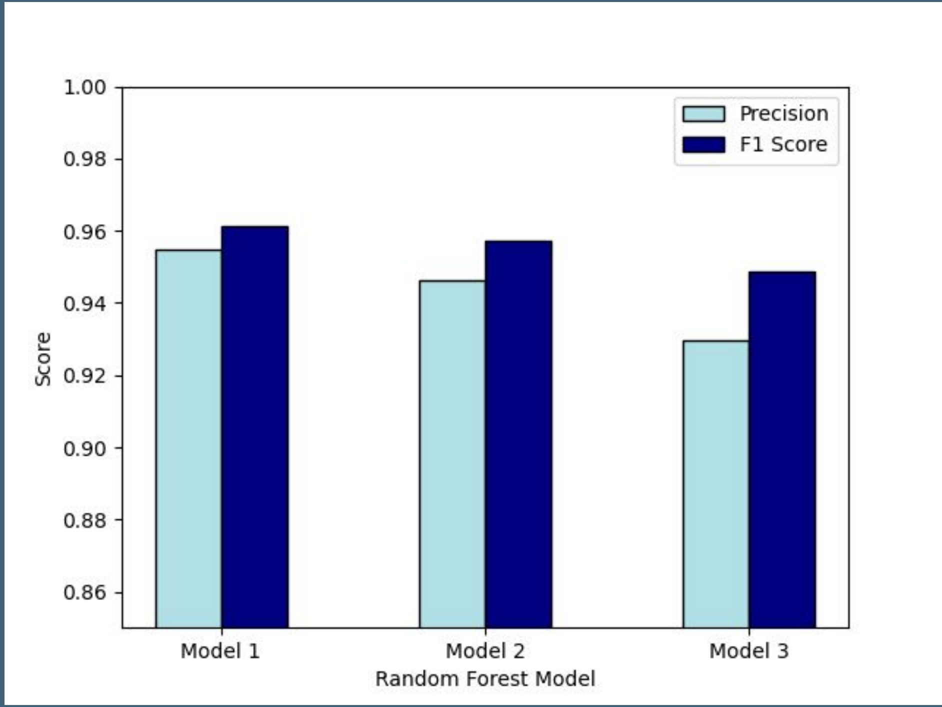
# Binary Classification Models

- Training/testing data for UCDs are from the UltracoolSheet
- “Contaminants” are queried to represent a standard sample of sky
- 864 UCDs (label: 1) and 796 Contaminants (label: 0) form the training set

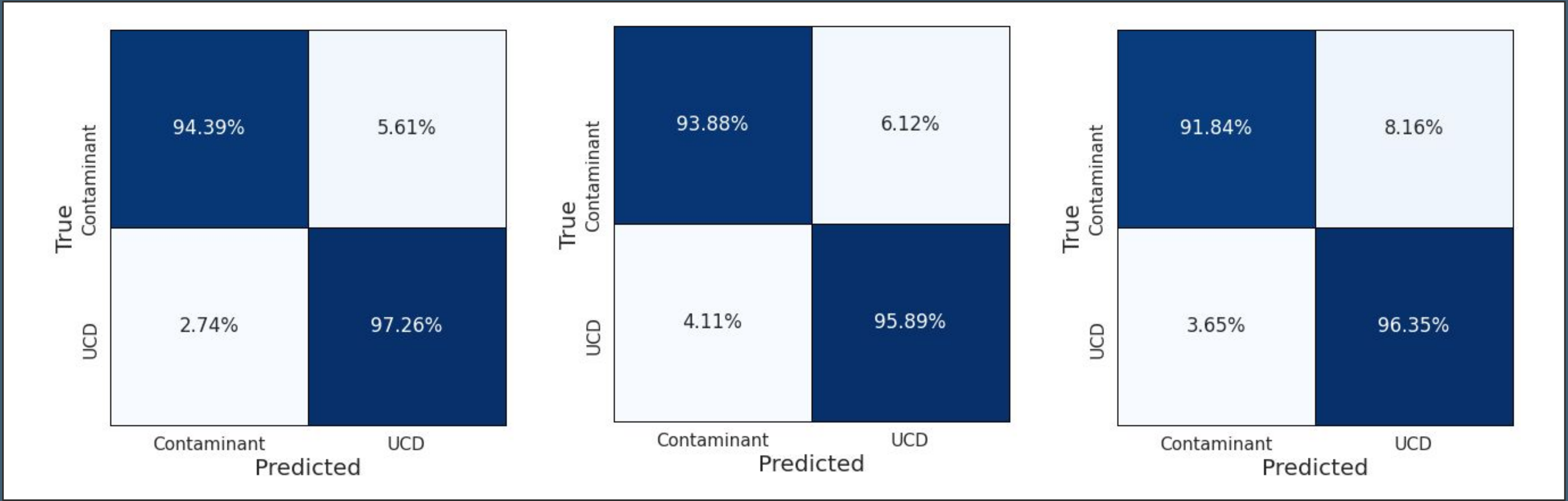
| **Model 1** | **Model 2** | **Model 3** |
| --- | --- | --- |
| Uses optical and infrared (rizYJ$K_s$W1W2) magnitudes and proper motions | Removes proper motions, only has rizYJ$K_s$W1W2 magnitudes | Based only on color from r-i to W1-W2 |

# Results: Binary Models

- Slight dip in performance from Model 1 to Model 2 when proper motions are removed
- Proper motions were the least important feature - included as close UCDs have larger motions
- Training just on colors still performs well

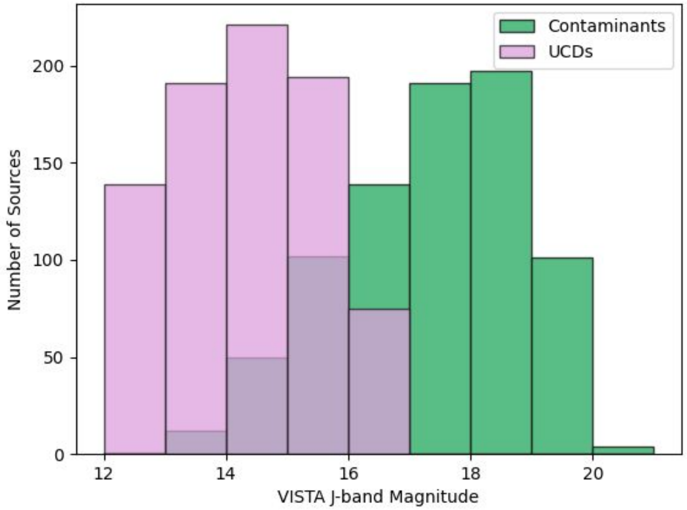


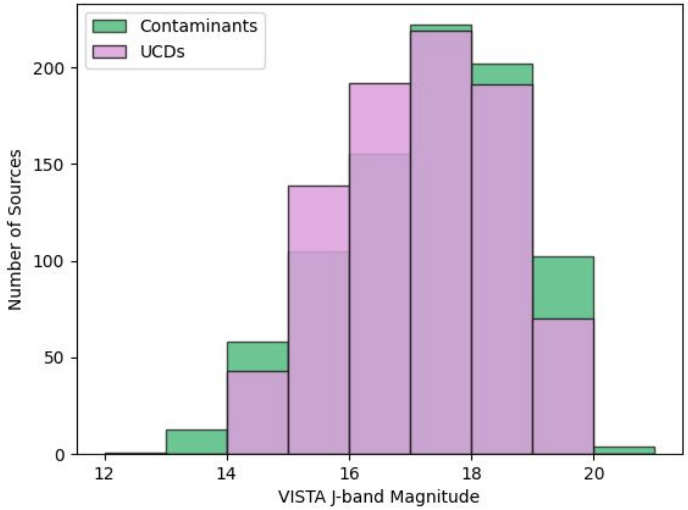
# Results: Binary Models




Model 1 | Model 2 | Model 3

# Brightness Bias

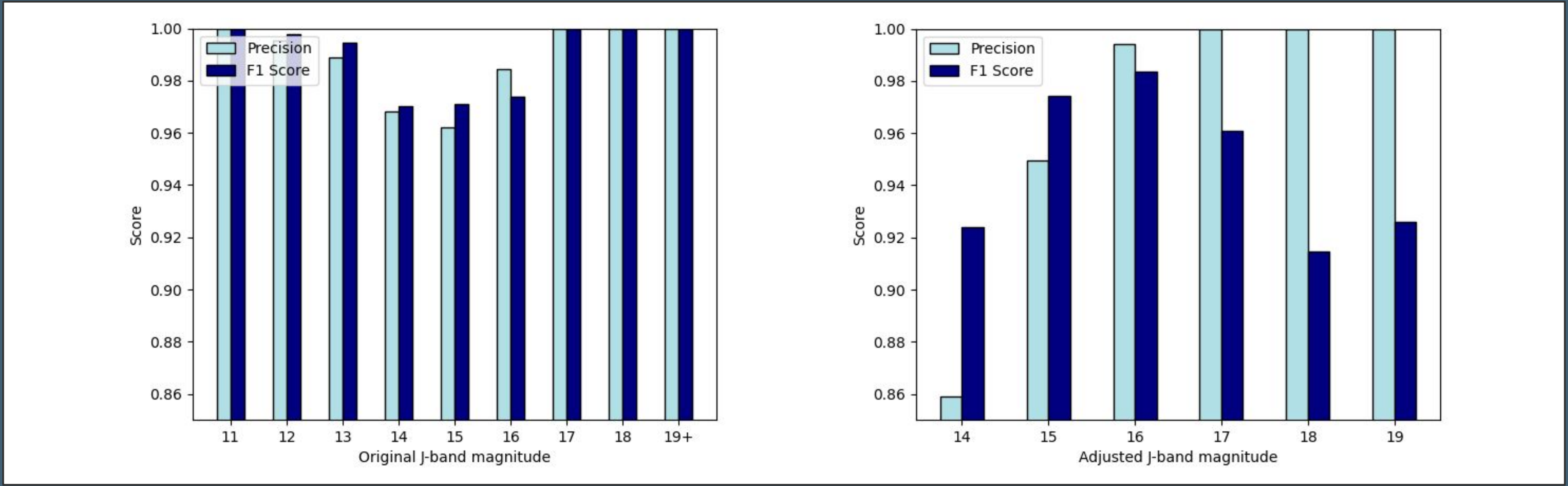



Adjusting VHS J-band photometry of sources to prevent brightness from biasing the classification

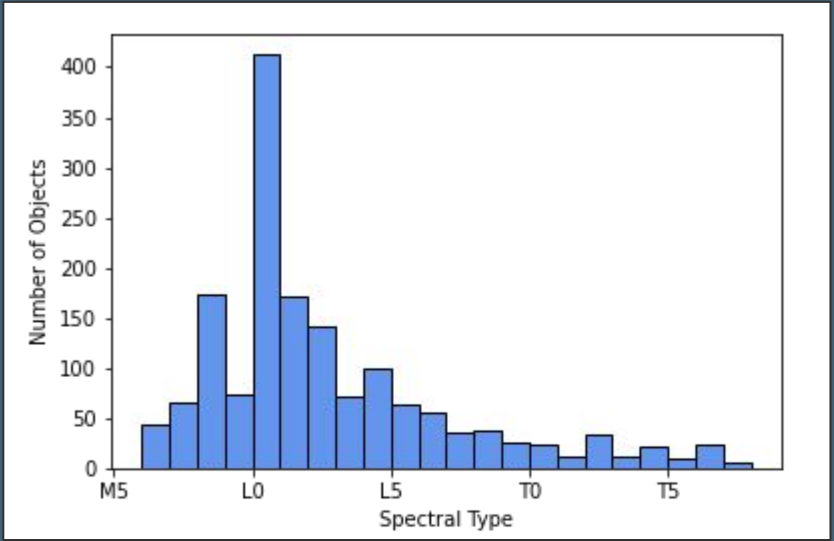
# Results: Magnitude Adjustment

- All sources in the training set are oversampled to 1000 using photometric uncertainty as noise for RF model testing
- Separated into bins 1 magnitude wide and tested through RF model 2



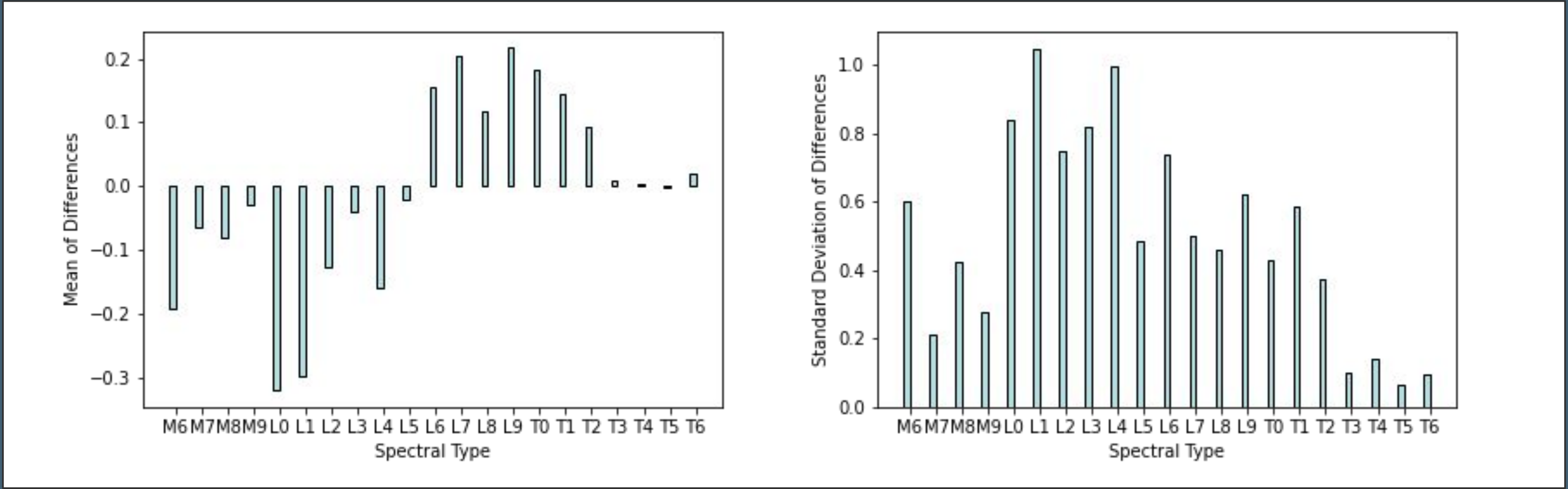
# Spectral Type

- The UltracoolSheet has very few T dwarfs with photometry, only 1 T9
- Oversampling with noise based on spectral type to increase T dwarf population
- Balanced all types to 1000 each



# Spectral Classification

- RF regression model trained on photometric colors from UltracoolSheet using z, Y, J, $K_s$, W1 and W2 bands
- Classification from M6 to T6



# Summary

- RF methods work accurately and quickly
- Still have limitations: need full photometry, potential for overfitting
- Broadly applicable to color-based/photometric selection of any type of source
- Can aid in creating a robust method of classification

# Future Work

- Further contaminant search of possible galactic sources
- RF parameter optimization
- Search for candidates in DES and running sources through our RF models
- Use population simulation models to measure Milky Way structure parameters

# Acknowledgements

UC San Diego

CASS

I would like to thank Professor Adam Burgasser and Dr. Chris Theissen for their direct input into this project, as well as the Cool Star Lab. This work relies on a large set of data made possible through Astro Data Lab and their query service.

Dr. Adam Burgasser

Dr. Chris Theissen

ASTRO DATALAB

Cool Star Lab

# References

Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20, doi: 10.3847/1538-4365/ac00b3

Aganze, C., Burgasser, A. J., Malkan, M., et al. 2022a, ApJ, 934, 73, doi: 10.3847/1538-4357/ac7053

—. 2022b, ApJ, 924, 114, doi: 10.3847/1538-4357/ac35ea

Best, W. M. J., Dupuy, T. J., Liu, M. C., Siverd, R. J., & Zhang, Z. 2020, The UltracoolSheet: Photometry, Astrometry, Spectroscopy, and Multiplicity for 3000+ Ultracool Dwarfs and Imaged Exoplanets, 1.0.0, Zenodo, doi: 10.5281/zenodo.4169085

Breiman, L. 2001, Machine Learning, 45, 5, doi: 10.1023/A:1010933404324

Carnero Rosell, A., Santiago, B., dalPonte, M., et al. 2019, MNRAS, 489, 5301, doi: 10.1093/mnras/stz2398

Gong, Z., Nava Vega, A., Gauna Gutierrez, E., et al. 2022, Research Notes of the American Astronomical Society, 6, 74, doi: 10.3847/2515-5172/ac6521

Gutierrez, E. G., Maytorena, A. M., Gong, Z., et al. 2022, Research Notes of the American Astronomical Society, 6, 75, doi: 10.3847/2515-5172/ac6522

Joergens, V. 2014, in Astrophysics and Space Science Library, Vol. 401, 50 Years of Brown Dwarfs, ed. V. Joergens, 1, doi: 10.1007/978-3-319-01162-2_1

# References


Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599, doi: 10.1111/j.1365-2966.2007.12040.x

Marocco, F., Caselden, D., Meisner, A. M., et al. 2019, ApJ, 881, 17, doi: 10.3847/1538-4357/ab2bf0

McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2021, VizieR Online Data Catalog, II/367

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825, doi: 10.48550/arXiv.1201.0490

Ryan, R. E., Thorman, P. A., Yan, H., et al. 2011, ApJ, 739, 83, doi: 10.1088/0004-637X/739/2/83

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163, doi: 10.1086/498708

The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0510346

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868, doi: 10.1088/0004-6256/140/6/1868

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579, doi: 10.1086/301513

# Thanks!

# Questions?

Contact:
mmdesai@ucsd.edu