UC San Diego

Finding Ultracool Dwarfs in Deep HST-WFC3 Surveys with Machine Learning

Christian Aganze (caganze@ucsd.edu, UCSD), Adam Burgasser (UCSD), Christopher Theissen (UCSD), Chih-chun Hsu (UCSD), Russell Ryan Jr. (STSCI), Danella Bardalez-Gagliuffi (AMNH)

PURPOSE

Compare methods for selecting M, L, T-dwarfs in deep spectroscopic surveys

Ultracool Dwarfs (UCDs) are objects with masses (<0.1 Msun, Kirkpatrick et al. 2005), which includes low-mass stars and brown dwarfs

WISP & 3D-HST SURVEYS

The WFC3 Infrared Spectroscopic Parallel Survey (WISPS, Atek et al. 2010) and 3D-HST survey (Momcheva et al. 2016) look for high-z galaxies using low-resolution (R ~200) near-infrared (1.1-1.7 microns) for more than 250,000 objects

SPECTRAL INDICES & TRAINING SETS

We define spectral indices to identify and classify UCDs by measuring the ratio of median fluxes in two different regions tracing water and methane features

We defined selection criteria using a sample of 2029 low resolution M5-T9 spectra from the SpexPrism Library (SPL, Burgasser 2014). To these, we add set of known Galaxies in both surveys to be used as sample of contaminants

METHOD 1: INDEX-INDEX SPACE SELECTION

We measure these indices for all sources in 3D-HST & WISPS, and our training set. To select UCDs, we enclose the training set into rectangular boxes with vertices defined to maximize completeness (see evaluation metrics). We pick the boxes with lowest contamination as selection regions

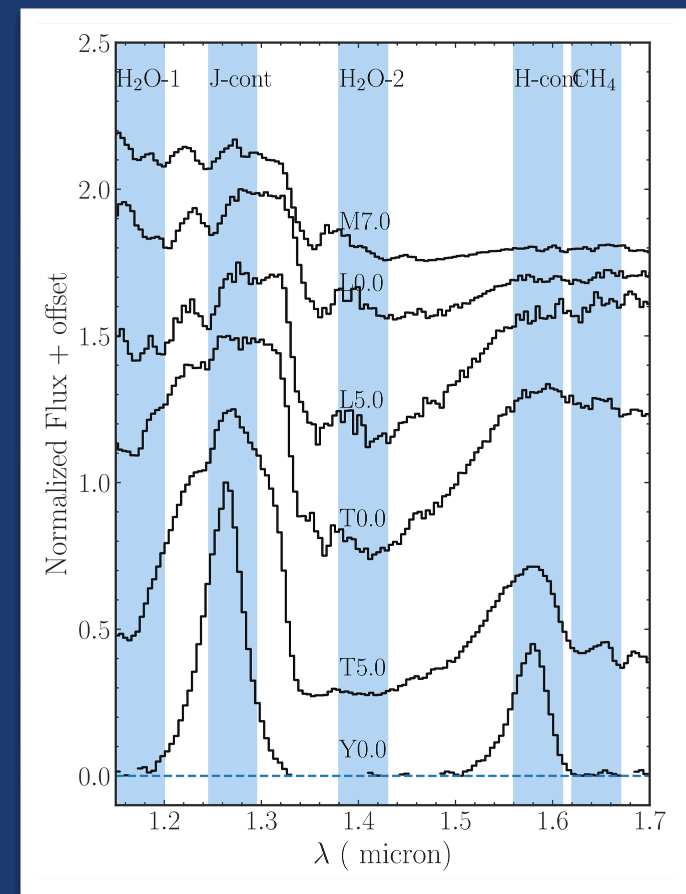

Fig 1: spectral standards showing regions used to defined indices

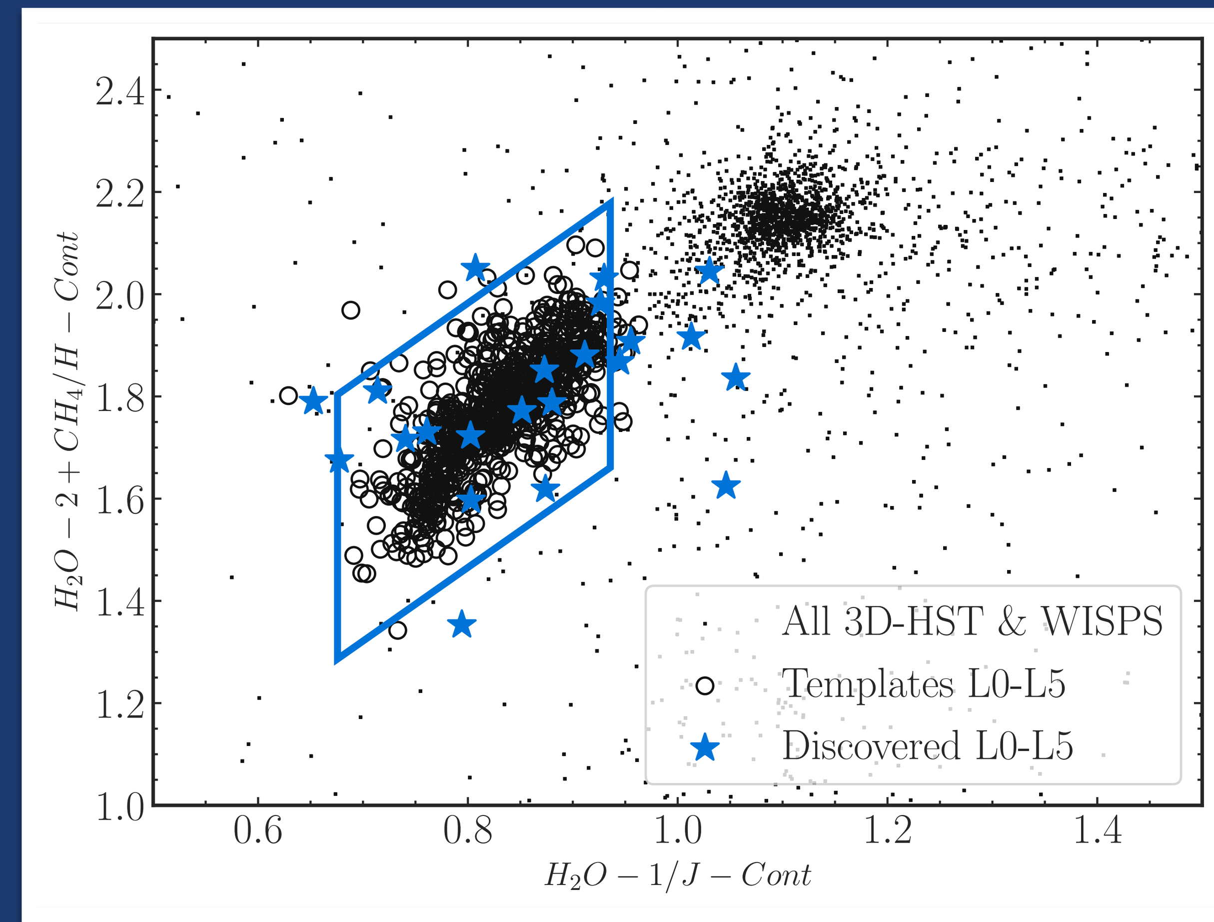

Fig 2: Demonstration of the index-selection method used to select L0-L5 dwarfs

METHOD 2: RANDOM FOREST

Labels: we trained a random forest model to identify objects into subgrouping: Galaxies, early M dwarfs (<M7), M7-L0, L and T

Features: in addition to spectral indices we added the S/N ratio measured in the J and H bands, the spectra type and goodness of fit statistics comparing each spectrum to a UCD spectral standard and a line

Main model parameters: 474 trees

Performance in training: we achieve a precision metrics of ~98% on the test set

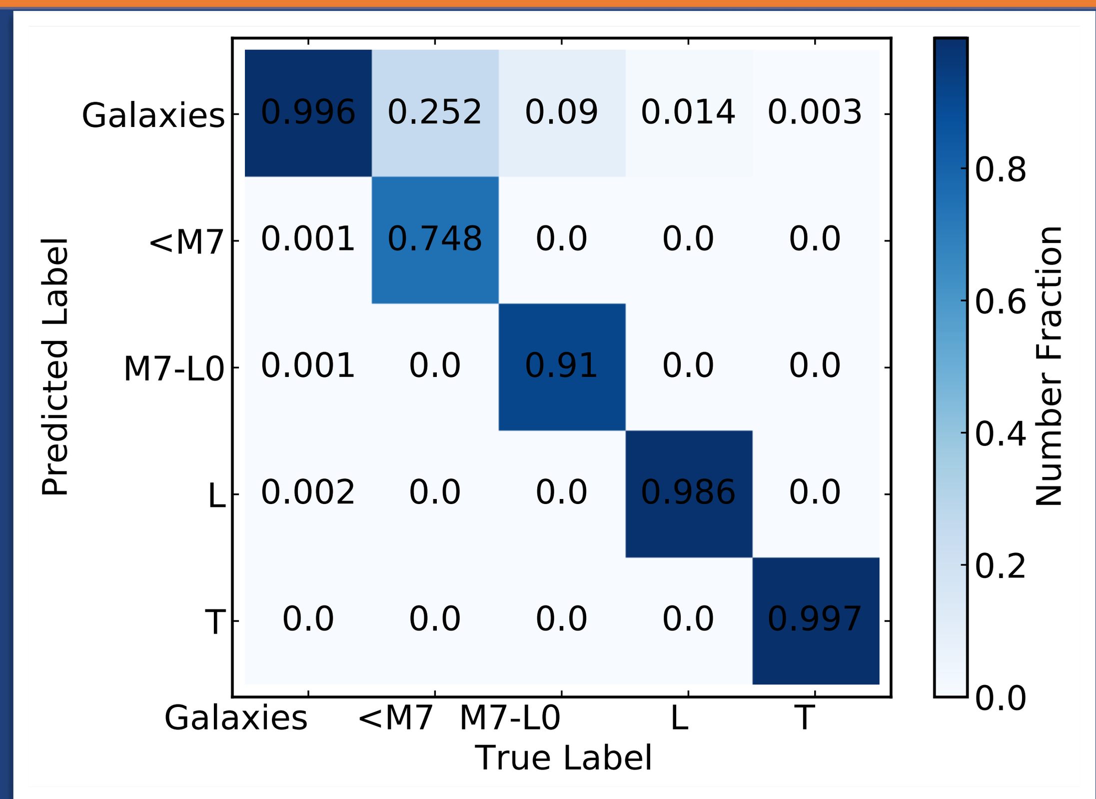

Fig 3: Confusion matrix for the test se showing the number of true positivists false positive

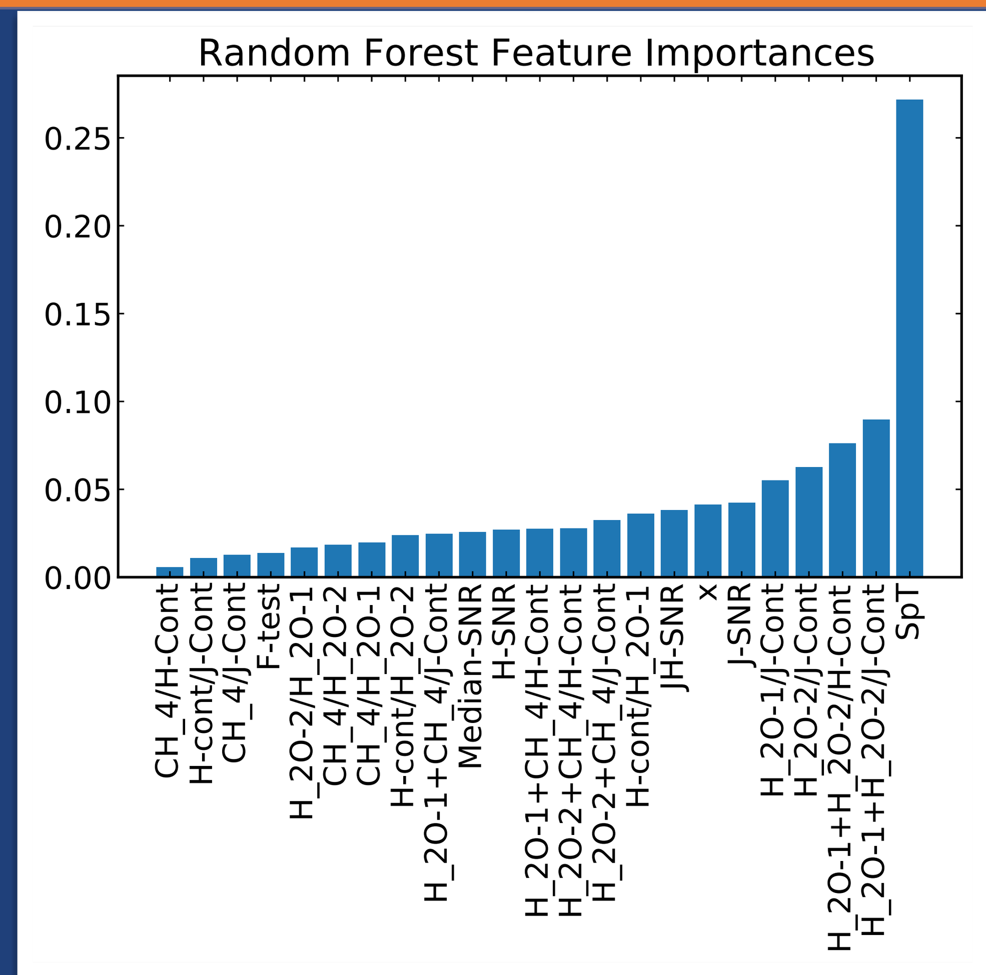

Fig 4: Feature importance

METHOD 3: ARTIFICIAL NEURAL NETWORK

Labels & Features: we used the same labels and features as the random forest method

Architectures: we explored two different architectures: a deep fully connected deep neural network (DNN) and a convolutional neural network (CNN). We optimized the number of parameters for these models using a random search

Main model parameters: 5 hidden layers for DNN, 1 convolutional layer and 3 hidden layer for CNN

Performance in training: precision of ~96 % for the DNN and ~95% for the CNN

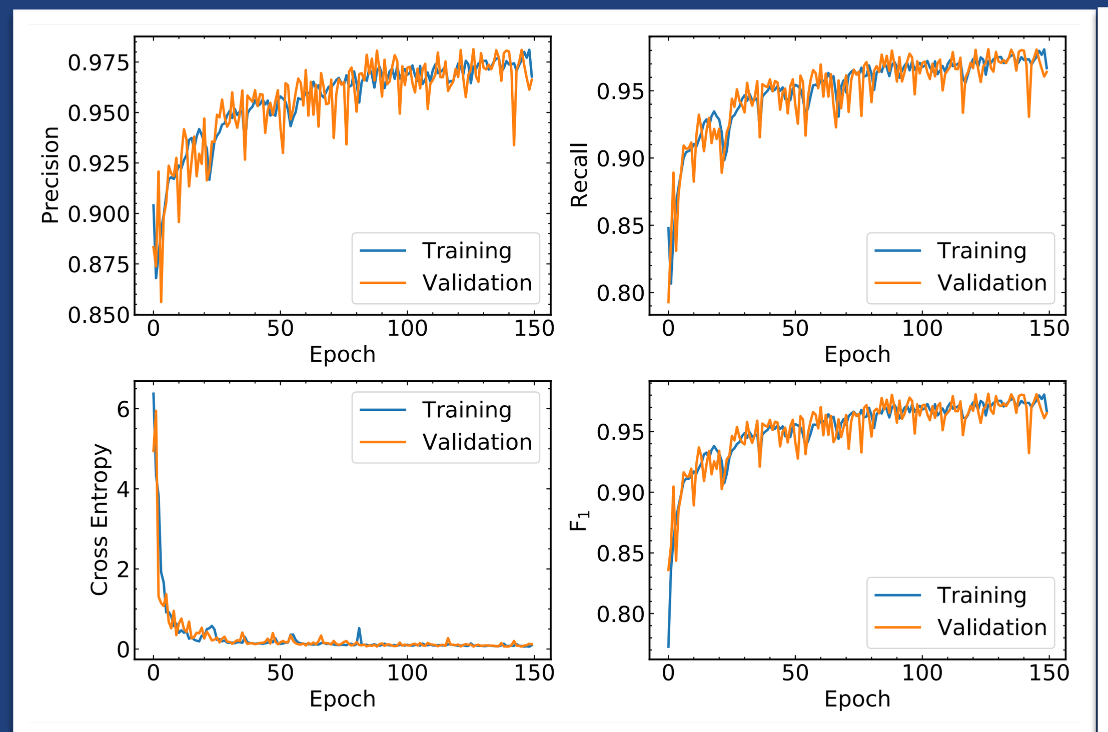

Fig 5: DNN performance metrics during training, using a batch size of 300

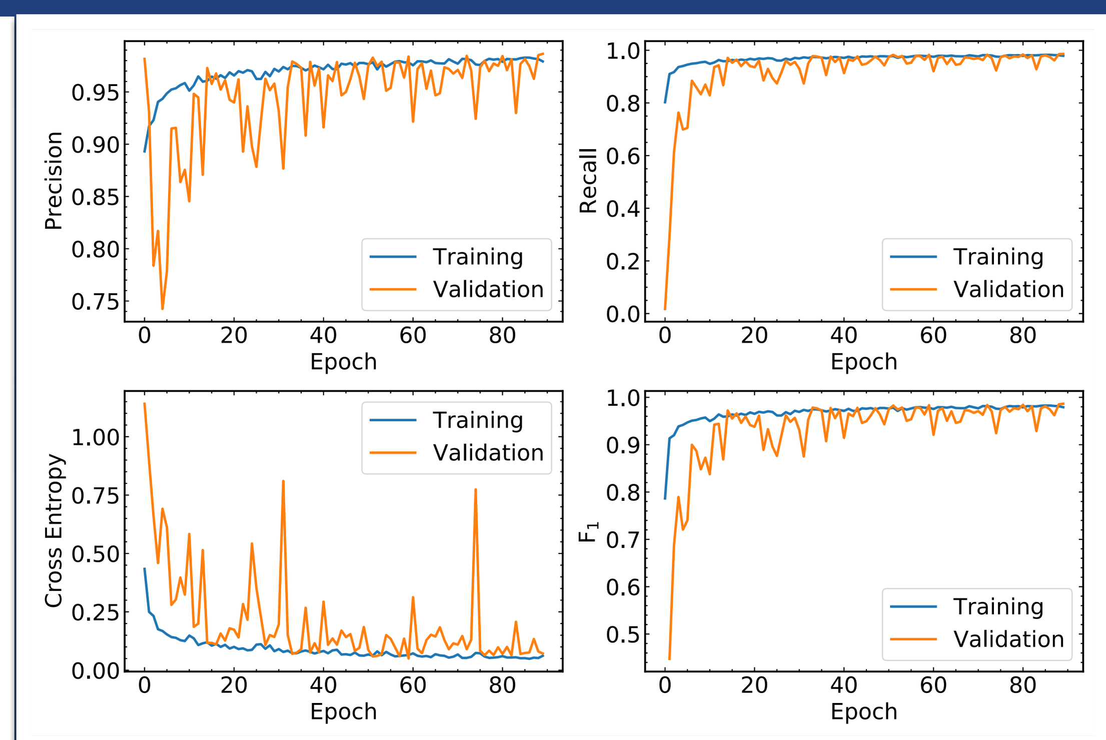

Fig 6: CNN performance metrics during training, using a batch size of 300

RESULTS: COMPARISON BETWEEN METHODS

Method	# UCDs	# Contaminants
Indices	152	1061
Random Forest	181	45
DNN	176	568
CNN	129	3070

The random forest outperforms traditional index selection and neural networks in the number of correctly identified UCDs and the least amount of contaminants, using the same features

EXAMPLE SPECTRA OF SELECTED UCDs

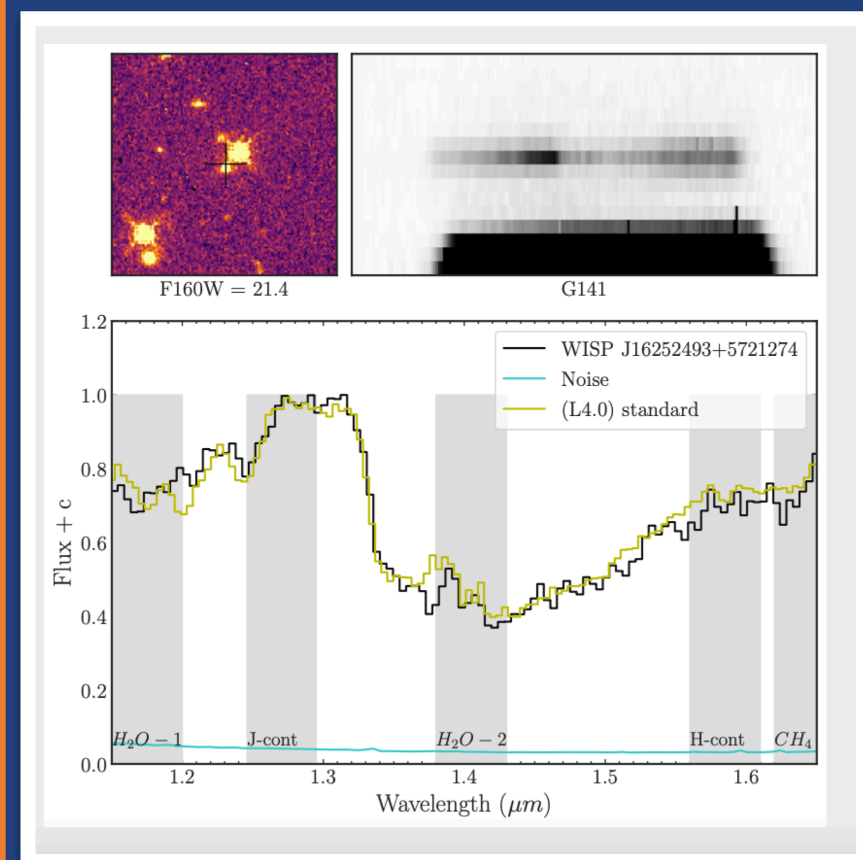

Fig 7: Spectrum of an L dwarf in the WISP Survey, next to brighter objects

Top: F160W image and 2D G141 spectrum

Bottom: 1D spectrum (black) compared to the best-fit spectral standard (yellow). The bands show regions used to define spectral indices

REFERENCES

Atek, H. et al. 2010, ApJ, 723, 104

Momcheva, I. et al. 2016, ApJS, 225, 27

Burgasser, A. J. 2014, in ASICS, Vol. 11, 7–16

Kirkpatrick, J. 2005, ARAA, 43, 195