Using the *SWS* Atlas dataset to predict *Spitzer/IRS* object types

Feb. 26, 2019

Slapdashery by Matt Shannon

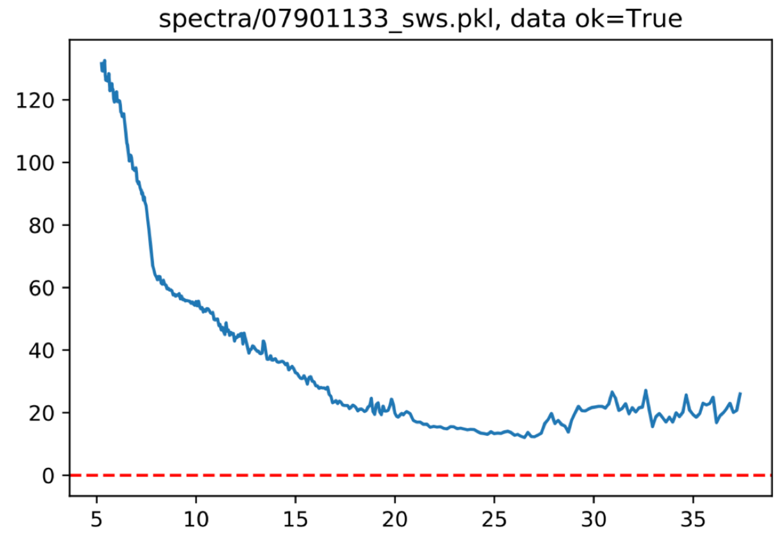
SWS Atlas – training set

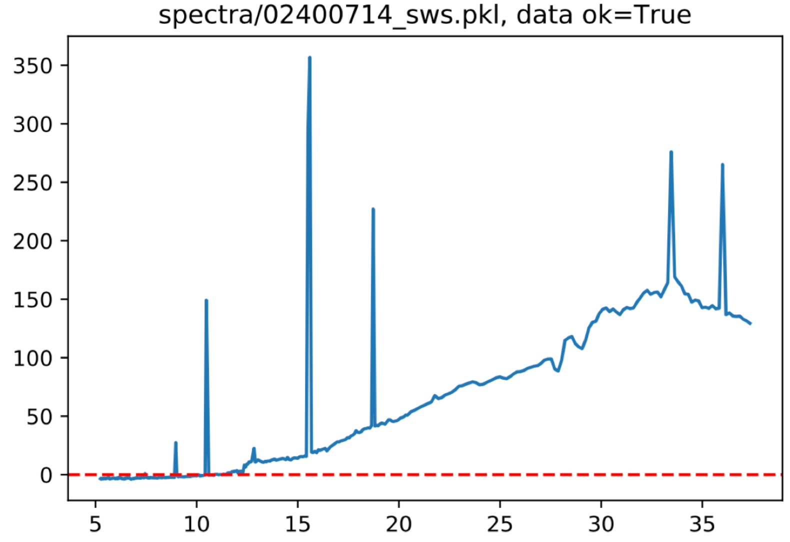
- 1239 spectra, Sloan+ 2003, 2.5-45.4 μm
  http://adsabs.harvard.edu/abs/2003ApJS..147..379S
- Mixture of object types

```
array(['', '**', '*inCl', '*inNeb', 'AGB*', 'Ae*', 'BYDra', 'Be*',
       'BlueCompG', 'BlueSG*', 'BrNeb', 'C*', 'CH', 'Candidate_AGB*',
       'Candidate_LP*', 'Candidate_post-AGB*', 'Cl*', 'ClG', 'Cloud',
       'ComGlob', 'DkNeb', 'EB*', 'EB*Algol', 'EB*betLyr', 'Em*',
       'Erupt*RCrB', 'FUOr', 'Flare*', 'Galaxy', 'GlCl', 'HH', 'HII',
       'IG', 'IR', 'Irregular_V*', 'LINER', 'LMXB', 'LPV*', 'Maser',
       'Mira', 'MolCld', 'Nova', 'Nova-like', 'OH/IR', 'OpCl', 'Orion_V*',
       'PM*', 'PN', 'PN?', 'PulsV*', 'PulsV*RVTau', 'PulsV*bCep',
       'PulsV*delSct', 'Pulsar', 'RGB*', 'RRLyr', 'RSCVn', 'RedSG*',
       'RfNeb', 'S*', 'SB*', 'SFregion', 'SG*', 'SN', 'SNR', 'Seyfert',
       'Seyfert_1', 'Seyfert_2', 'Star', 'Symbiotic*', 'TTau*', 'V*',
       'V*?', 'WR*', 'YSO', 'YellowSG*', 'deltaCep', 'denseCore',
       'gammaBurst', 'multiple_object', 'pMS*', 'post-AGB*', 'semi-regV*'],
      dtype=object)
```

SWS Atlas – training set

- Metadata:
  - Object type
    - From SIMBAD (not always reliable)
  - Primary classifier (group)
    1. Naked stars
    2. Stars with dust
    3. Warm, dusty objects
    4. Cool, dusty objects
    5. Very red objects
    6. Continuum-free objects but having emission lines
    7. Flux-free and/or fatally flawed spectra





SWS Atlas – training set

- Metadata:
  - Secondary classifiers
  - Qualifiers

TABLE 3
LEVEL 2 SUFFIXES

Suffix	Description
e	Emission lines (e.g., H recombination, atomic fine structure)
u	UIR features present, but not dominant feature
p	Fits in given category but is peculiar
:, ::	Uncertain (either noisy or odd)

TABLE 2
LEVEL 2 CLASSIFICATION DEFINITIONS

Class	Description
SE	Silicate (or oxygen-rich) dust emission (10–12 and 18–20 μm)
SB	Silicate emission in self-absorption (10 μm)
SA.........	Silicate absorption (10–12 μm)
SC.........	Silicate emission from crystalline grains (33, 40, 43 μm)
SEC	Silicate emission from crystalline grains (11, 19, 23, 33 μm)
CE	Carbon-rich dust emission, primarily from SiC (11.5 μm)
CR	Carbon-rich dust emission in a reddened shell (with features at 11.5 and 26 μm, often 13.7 μm absorption)
CT	8, 11.5, 21, 26 μm, no 13.7 absorption
CN........	Carbon-rich nebulae
C/SE	Carbon-rich, plus silicate emission (10–12 μm)
C/SC.....	Carbon-rich, plus crystalline silicate emission
U/SC.....	Crystalline silicate and UIR emission features
U	Prominent UIR emission features
PN	Many prominent atomic fine-structure lines typical of PNs
PU	As PN, but with strong UIR emission
W.........	Emission peaks 6–8 μm
F..........	Basically featureless
E..........	Strong emission lines
M.........	Miscellaneous

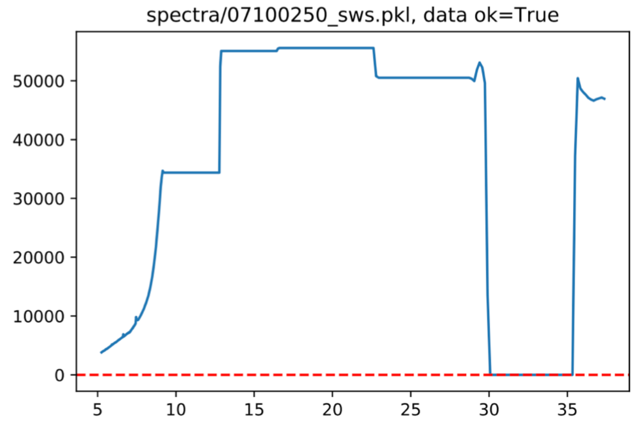
Data processing

Initial steps

	wavelength	flux	spec_error	norm_error
count	48924.000000	48924.000000	48924.000000	48924.000000
mean	13.890253	29.865271	1.316019	1.436715
std	11.323857	77.273079	2.070261	2.219514
min	2.360000	-6.720000	0.080000	0.080000
25%	4.717688	-2.600000	0.270000	0.390000
50%	10.250250	-0.660000	0.560000	0.600000
75%	19.009378	43.250000	0.990000	1.000000
max	45.389999	3170.389893	16.940001	17.129999

- Convert spectra to Pandas DataFrame objects
- Store metadata
- Exclude bad spectra

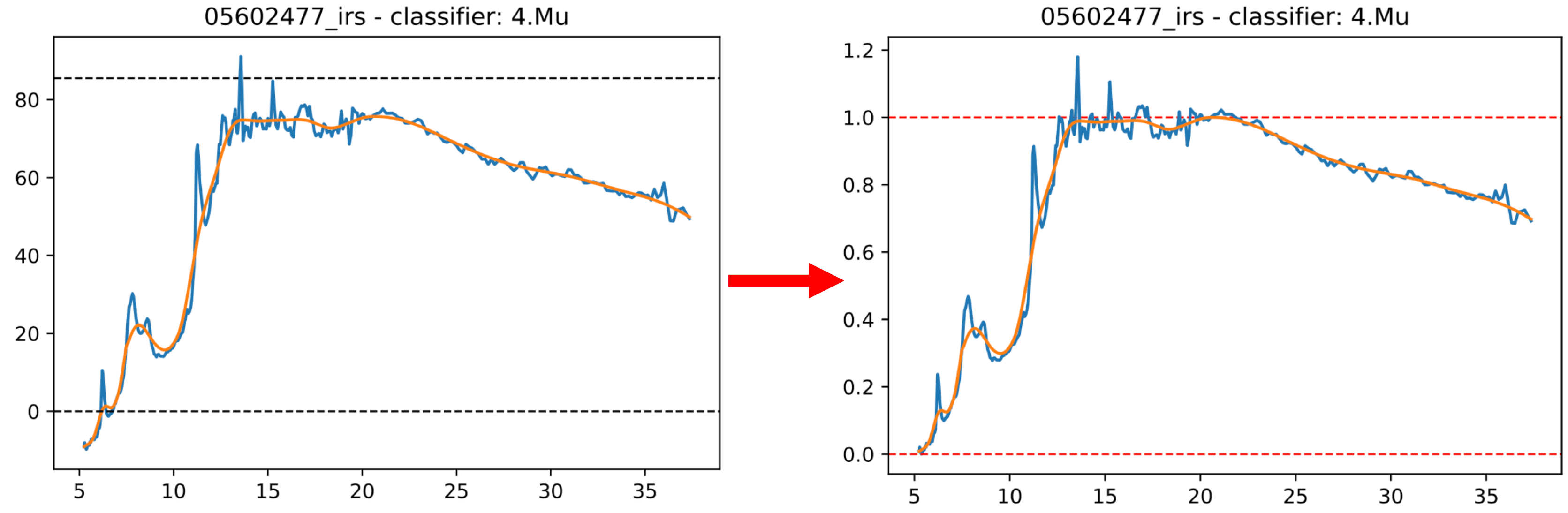
	object_name	tdt	ra	dec	full_classifier	group	subgroup	uncertainty_flag	note	file_path	object_type	data_ok
0	NGC 6543	2400714	269.639167	66.633194	4.PN	4	PN			spectra/02400714_irs.pkl	PN	True
1	NGC 6543	2400807	269.639167	66.633194	4.PN	4	PN			spectra/02400807_irs.pkl	PN	True
2	NGC 6543	2400910	269.639125	66.633194	4.PN	4	PN			spectra/02400910_irs.pkl	PN	True
3	NGC 7027	2401183	316.757125	42.235861	4.PU	4	PU			spectra/02401183_irs.pkl	PN	True
4	{gamma} Dra	2401579	269.151708	51.488972	1.NO	1	NO		(0,0)	spectra/02401579_irs.pkl		True



Remaining: 741 of 1239 spectra

Normalize & regrid

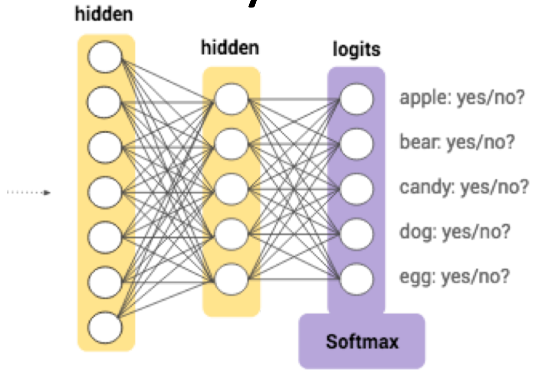
- ML algorithms like it when data ∈ {0, 1} (ish)
  - Smooth data to determine min/max, shift & scale; downsample to *IRS/SL-LL* resolution
- Some data starts below zero due to SWS stitching



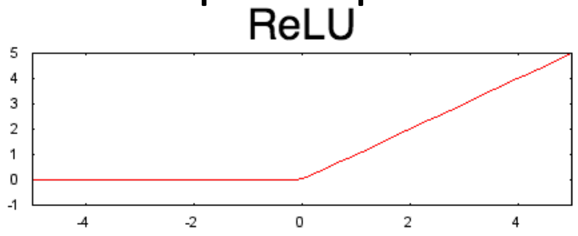
SWSnet model

Model parameters

- Output layer: 5 groups (naked, dusty, etc.), normalized probability

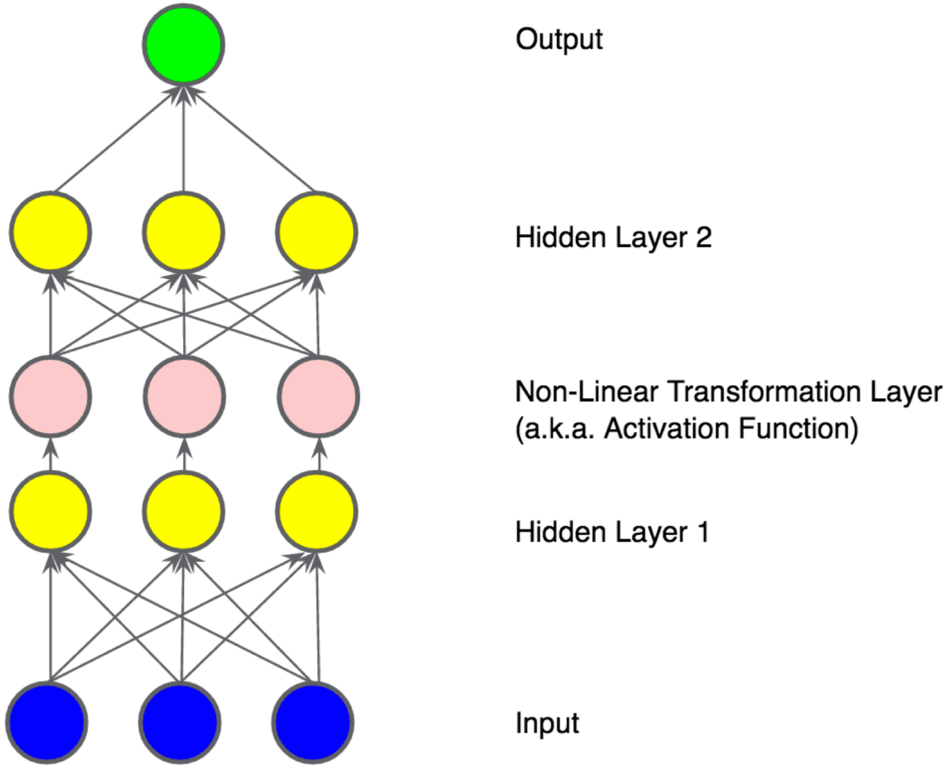


- Two hidden layers:
  - 64 neurons each, interconnected
  - Transform output of previous neurons via:



- Input layer: 741 spectra (359 resolution elements each)

```
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 64)                23040
_________________________________________________________________
dense_1 (Dense)              (None, 64)                4160
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 325
=================================================================
Total params: 27,525
Trainable params: 27,525
Non-trainable params: 0
_________________________________________________________________
```

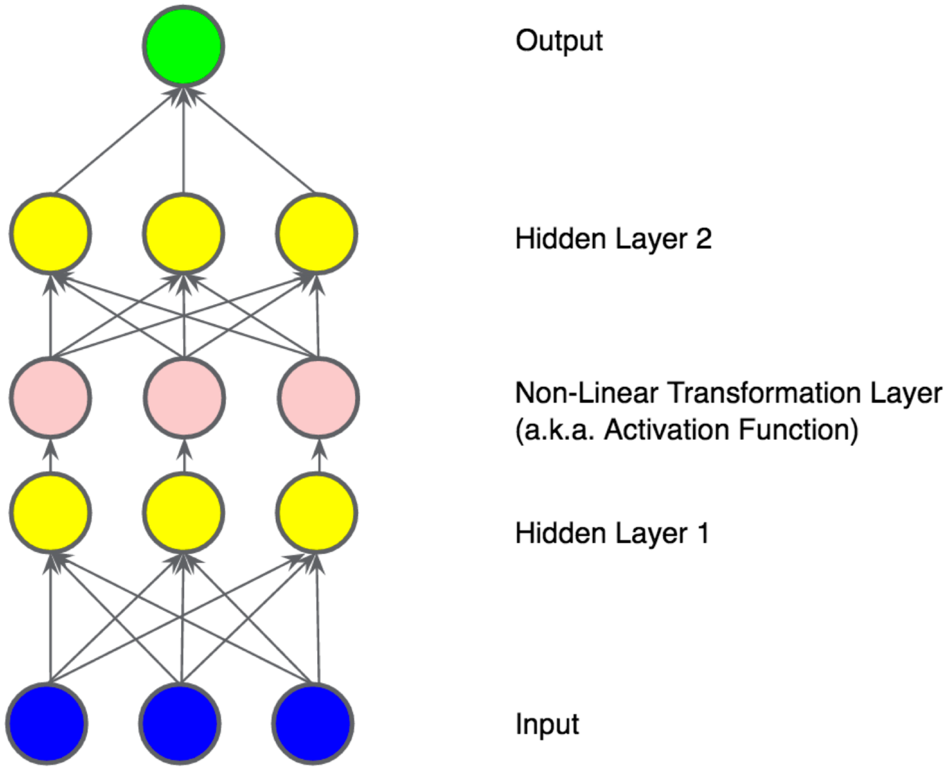


Model parameters

- Limit overfitting via:
  - Early stopping (stop iterating if test accuracy is not improving after 5 iterations)
  - Penalize high coefficients by L2-norm, i.e. coefficients^2
- Adjust weights with Adam optimizer (variable learning rates per neuron, etc.)

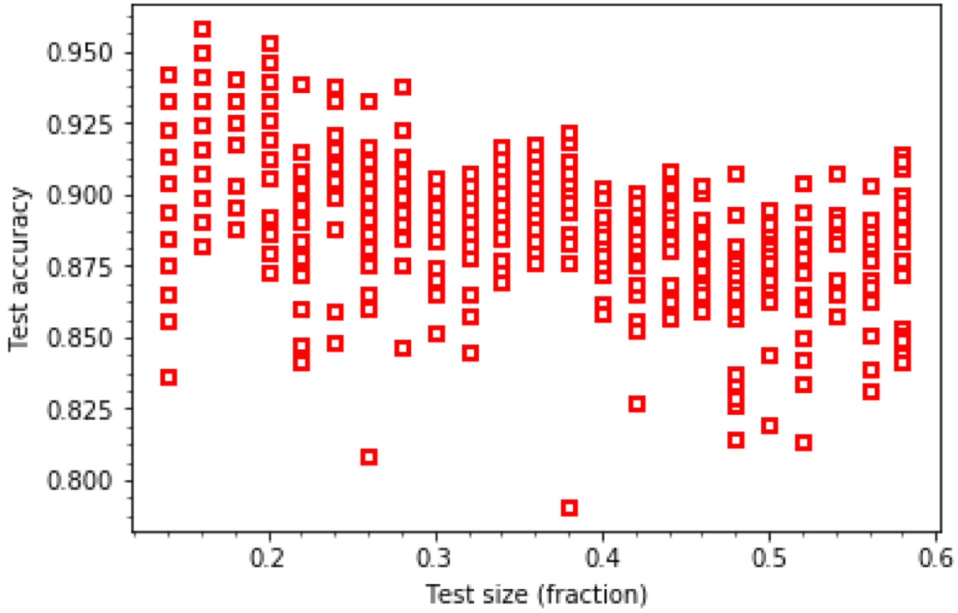
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	23040
dense_1 (Dense)	(None, 64)	4160
dense_2 (Dense)	(None, 5)	325

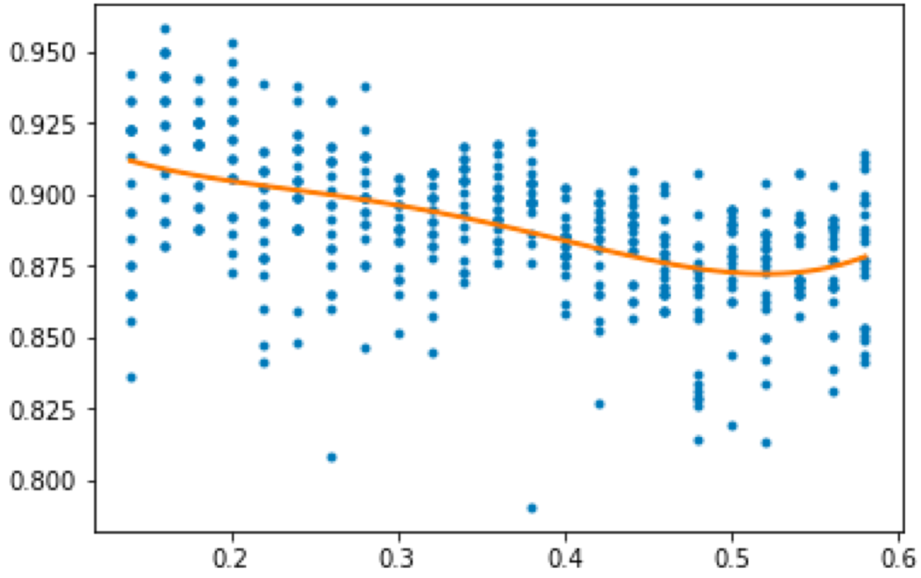
Total params: 27,525
Trainable params: 27,525
Non-trainable params: 0



Model results/validation

- Adopted test size = 0.35, training size = 0.65 (of 741 spectra)
- ~88-90% accuracy on SWS dataset for primary ‘Group’ classification
  - Naked, dusty, etc.
  - Mostly probes SED

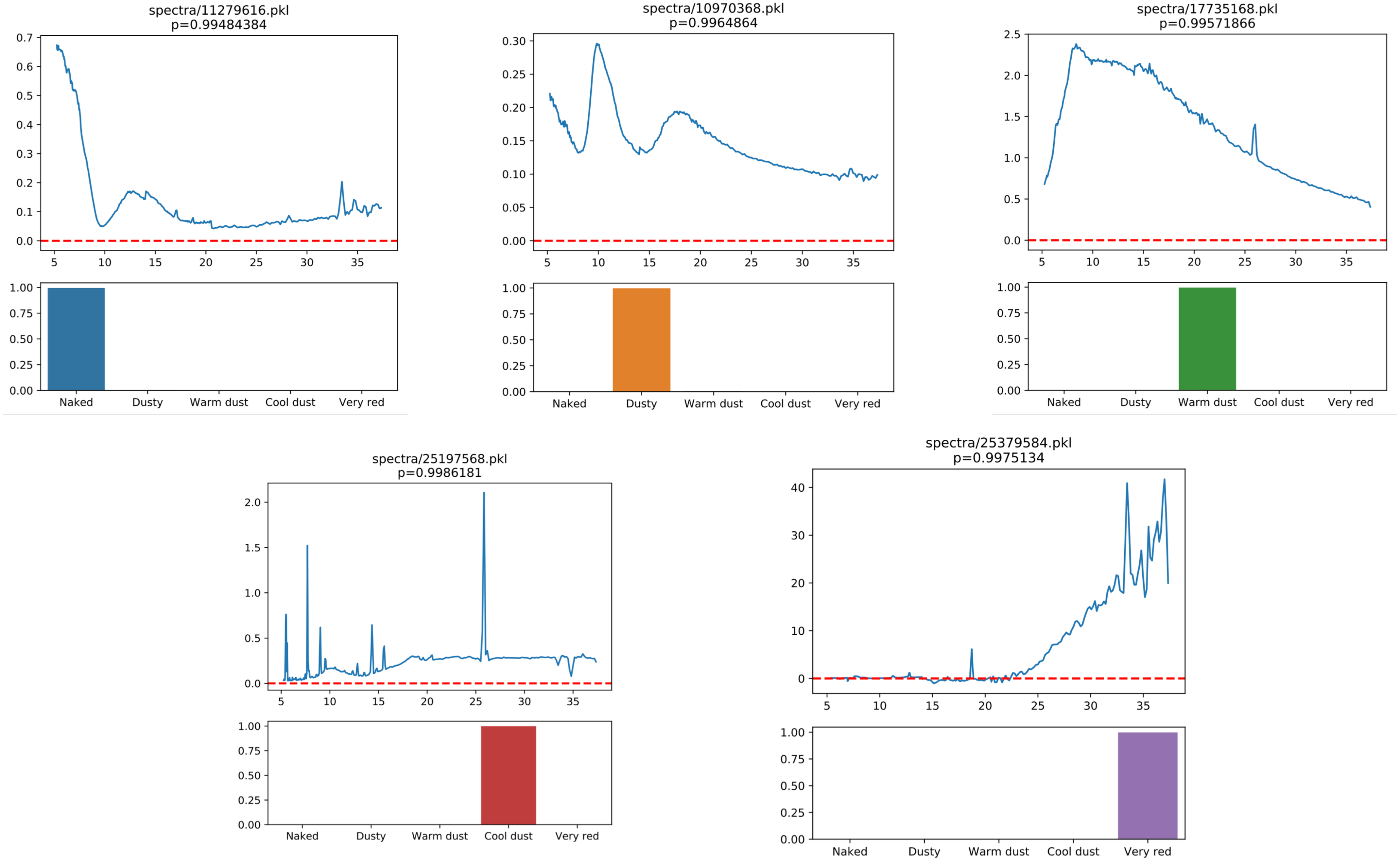




Applying the model to *Spitzer* spectra

6733 spectra in total.

See laptop for fliipbook, but a few examples (groups 1-5):


spectra/11279616.pkl
p=0.99484384
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
5
10
15
20
25
30
35
1.00
0.75
0.50
0.25
0.00
Naked
Dusty
Warm dust
Cool dust
Very red
spectra/10970368.pkl
p=0.9964864
0.30
0.25
0.20
0.15
0.10
0.05
0.00
5
10
15
20
25
30
35
1.00
0.75
0.50
0.25
0.00
Naked
Dusty
Warm dust
Cool dust
Very red
spectra/17735168.pkl
p=0.99571866
2.5
2.0
1.5
1.0
0.5
0.0
5
10
15
20
25
30
35
1.00
0.75
0.50
0.25
0.00
Naked
Dusty
Warm dust
Cool dust
Very red
spectra/25197568.pkl
p=0.9986181
2.0
1.5
1.0
0.5
0.0
5
10
15
20
25
30
35
1.00
0.75
0.50
0.25
0.00
Naked
Dusty
Warm dust
Cool dust
Very red
spectra/25379584.pkl
p=0.9975134
40
30
20
10
0
5
10
15
20
25
30
35
1.00
0.75
0.50
0.25
0.00
Naked
Dusty
Warm dust
Cool dust
Very red

Results, primary classifier

(80%+ probability in a single category)

1. Naked stars – 1816 sources
2. Dusty stars – 330 sources
3. Warm dust – 64 sources
4. Cool dust – 1015 sources
5. Very red – 1724 sources

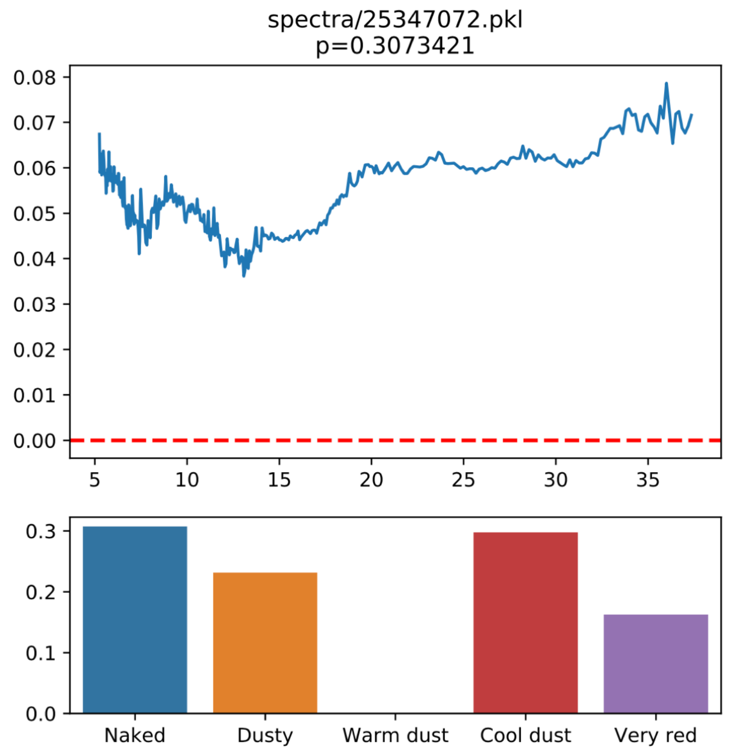
Those with 70-80% certainty – 572

Those with 60-70% certainty – 510

Those with 50-60% certainty – 492

Less than 50% -- 164

Worst case scenario:



Possible next steps

- Utilize secondary classifiers!
- Train within a single group instead? Use separate NN to determine subclassification?
- Train for presence of UIRs?

TABLE 3

LEVEL 2 SUFFIXES

Suffix	Description
e.........	Emission lines (e.g., H recombination, atomic fine structure)
u.........	UIR features present, but not dominant feature
p	Fits in given category but is peculiar
:, ::	Uncertain (either noisy or odd)

TABLE 2

LEVEL 2 CLASSIFICATION DEFINITIONS

Class	Description
SE.........	Silicate (or oxygen-rich) dust emission (10–12 and 18–20 μm)
SB.........	Silicate emission in self-absorption (10 μm)
SA.........	Silicate absorption (10–12 μm)
SC.........	Silicate emission from crystalline grains (33, 40, 43 μm)
SEC	Silicate emission from crystalline grains (11, 19, 23, 33 μm)
CE	Carbon-rich dust emission, primarily from SiC (11.5 μm)
CR	Carbon-rich dust emission in a reddened shell (with features at 11.5 and 26 μm, often 13.7 μm absorption)
CT	8, 11.5, 21, 26 μm, no 13.7 absorption
CN........	Carbon-rich nebulae
C/SE	Carbon-rich, plus silicate emission (10–12 μm)
C/SC.....	Carbon-rich, plus crystalline silicate emission
U/SC.....	Crystalline silicate and UIR emission features
U	Prominent UIR emission features
PN	Many prominent atomic fine-structure lines typical of PNs
PU	As PN, but with strong UIR emission
W..........	Emission peaks 6–8 μm
F...........	Basically featureless
E...........	Strong emission lines
M..........	Miscellaneous