Final Capstone – Galaxy Image Classification

Jeff Biehle

September 13, 2019

A little bit about me. . .

- Live in Dripping Springs, TX
  - Outside Austin, “just west of ‘Weird’ “
- 30+ years in high-tech
  - Primarily business development/strategic alliances
- Decided I needed a career switch
- Have been intrigued with data science for several years
- Joined Thinkful in February

Overview

Why Galaxies?

The question(s)

Data fun!

Model Selection/ Tuning

Conclusion/ Challenges

Why galaxies?

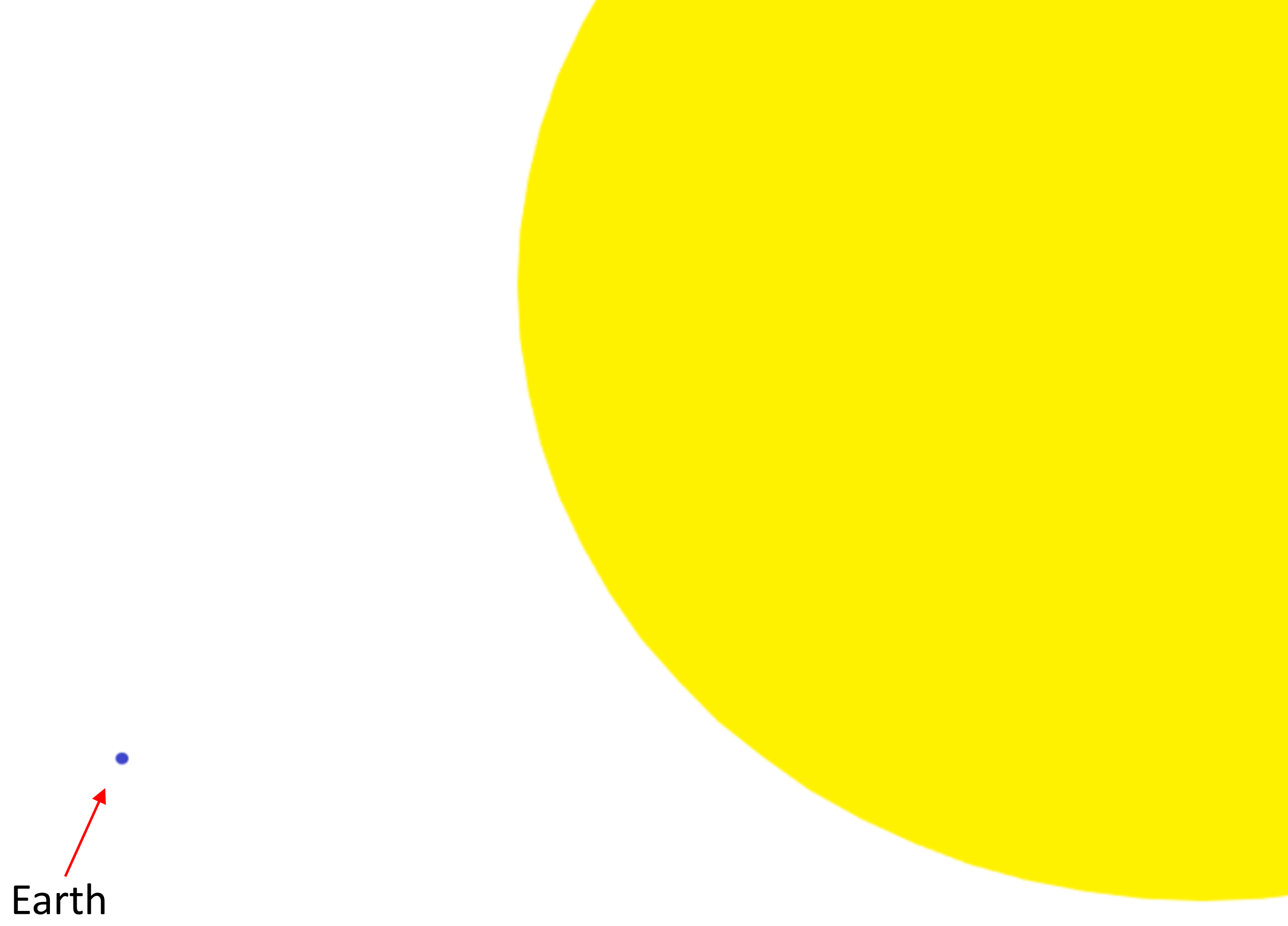
- 50th anniversary of Apollo 11, so I felt compelled to do something space-themed
- Space buff since I was a kid
  - Barely remember the landing
- I was a short-term member of the Galaxy Zoo project when first introduced 12 years ago

Pardon me while I geek out on some fun astronomy data

Some perspective on size

The universe is unimaginably large

- The sun
  - 865,000 miles in diameter
    - If the earth were the size of basketball, the sun would be over 85 feet tall
    - You could fit 1.3 million earths inside the sun
  - 93 million miles away
    - If you could pilot a jet to the sun at 600mph without stopping, it would take you nearly 18 years to reach it



Some perspective on size

Proxima Centauri

- The nearest star to our solar system
- 4.243 light years = 25,000,000,000,000 miles

- Voyager II travels at about 10 miles per second
  - Could fly from New York to LA in 4.5 minutes
  - It would take over 80,000 years to reach Proxima Centauri

Some perspective on size

The Milky Way

- 100,000 light years across
- Estimates range from 100 billion to 400 billion stars
- Nearest galaxy is Andromeda
    - 2.3 million light years away

Some perspective on size

The universe

- 93 billion light years across
- Each object/dot in this picture is a galaxy
- Estimated to contain 100 billion galaxies
- On average each galaxy probably contains 100 billion stars
- Thus there are (theoretically) 10^22 stars in the universe
- 10,000,000,000,000,000,000,000 (10 sextillion) stars

Hubble Telescope Deep Field Survey

Some perspective on size

The universe

- Although impossible to know for sure, it's generally believed in the scientific community that there are more stars in the universe than all the grains of sand on all Earth's beaches!

Hubble Telescope Deep Field Survey

The data set

Found on Kaggle

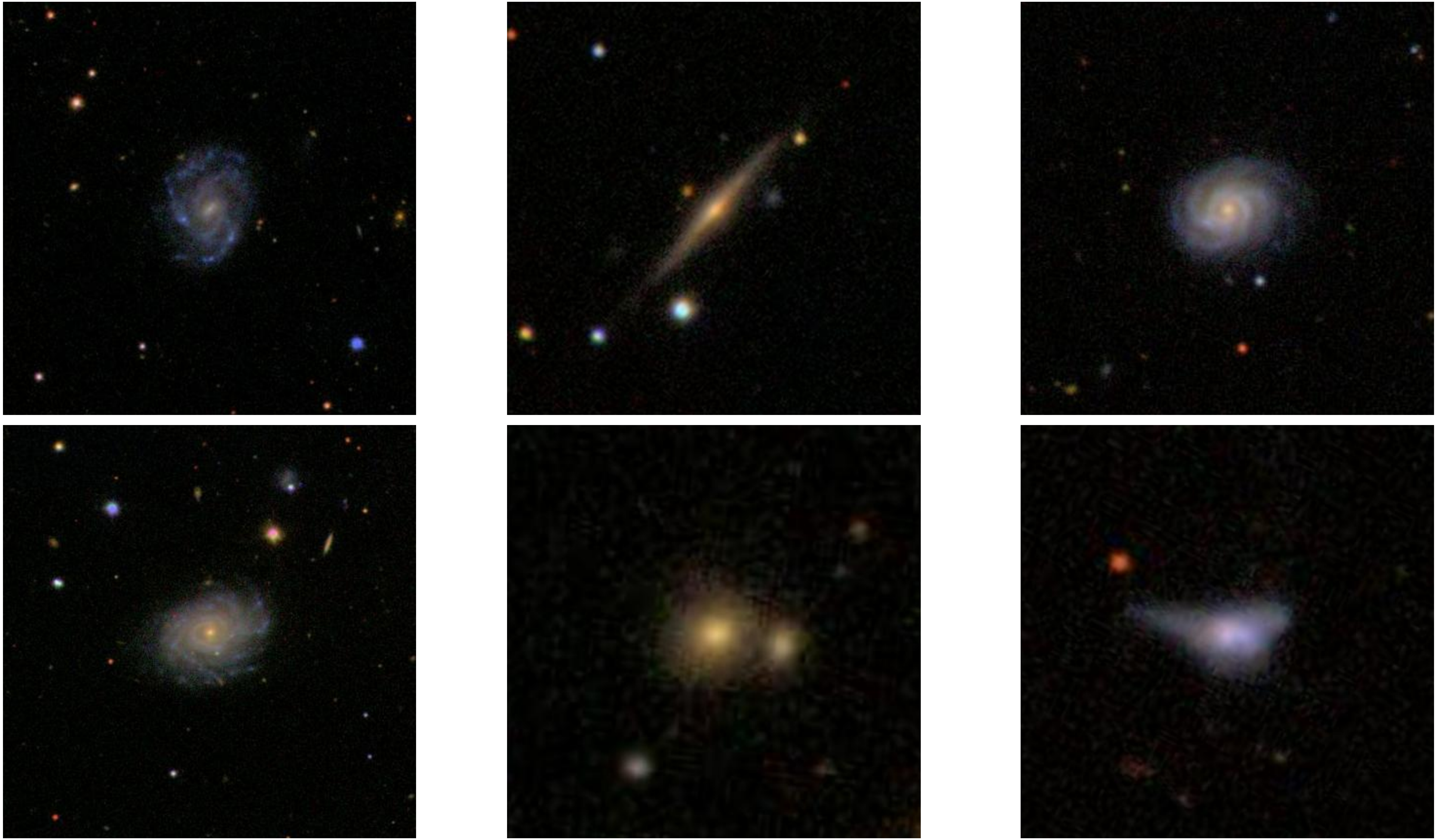
https://bit.ly/2J9XjLt

Sponsored by Galaxy Zoo Project

Over 60,000 galaxies from multiple years of sky surveys

The data set

- The Galaxy Zoo Project
  - Established in 2007
  - Thousands of volunteers to classify hundreds of thousands of galaxy images
  - Data is to help astronomers understand the distribution of types and features of galaxies



11 questions, 37 total responses

The question(s)

Q1. Is the object a smooth galaxy, a galaxy with features/disk or a star? *3 responses*
Q2. Is it edge-on? *2 responses*
Q3. Is there a bar? *2 responses*
Q4. Is there a spiral pattern? *2 responses*
Q5. How prominent is the central bulge? *4 responses*
Q6. Is there anything "odd" about the galaxy? *2 responses*
Q7. How round is the smooth galaxy? *3 responses*
Q8. What is the odd feature? *7 responses*
Q9. What shape is the bulge in the edge-on galaxy? *3 responses*
Q10. How tightly wound are the spiral arms? *3 responses*
Q11. How many spiral arms are there? *6 responses*

11 questions, 37 total responses

The question(s)

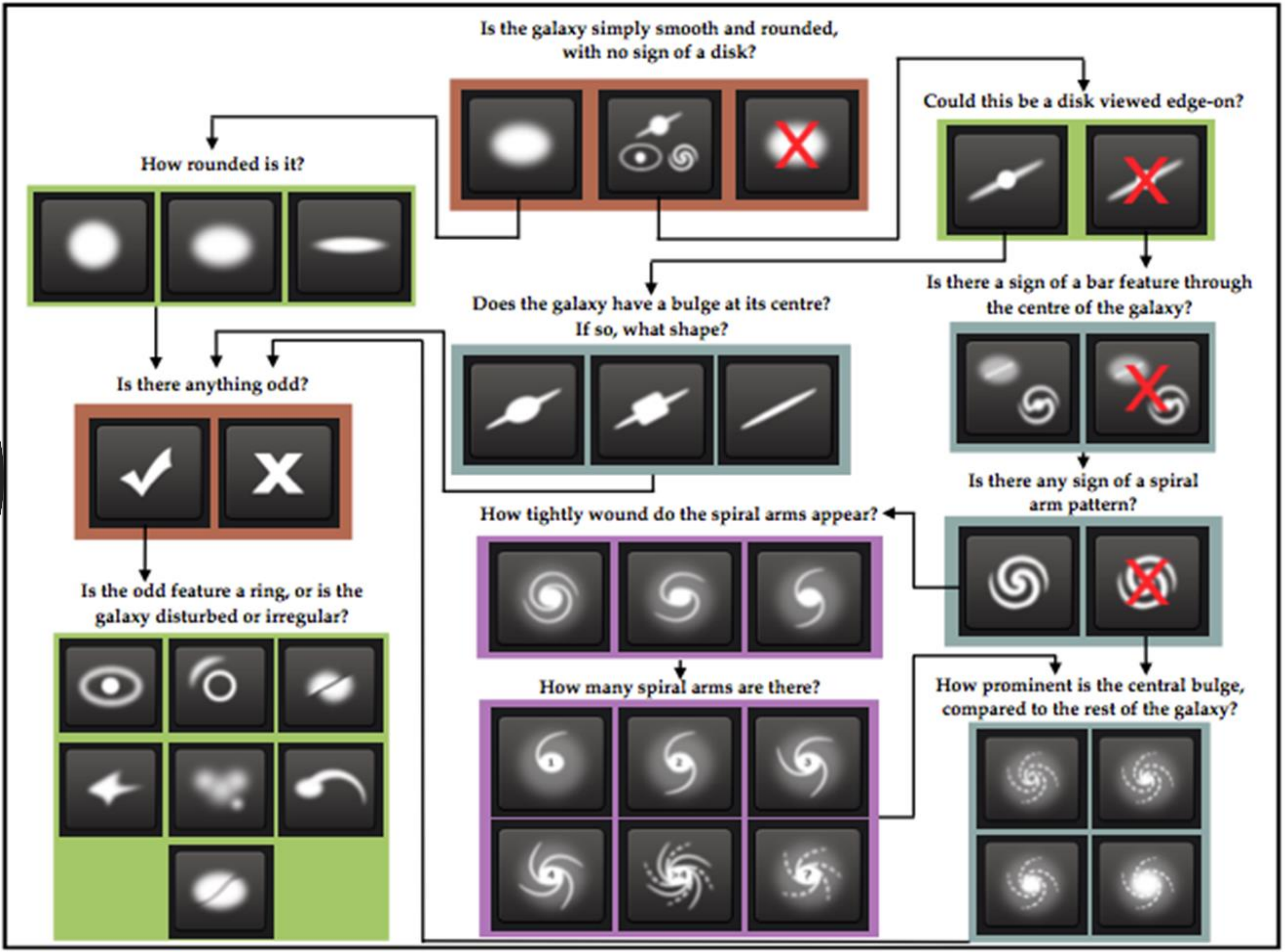
Q1. Is the object a smooth galaxy, a galaxy with features/disk or a star? *3 responses*

Q2. Is it edge-on? *2 responses*

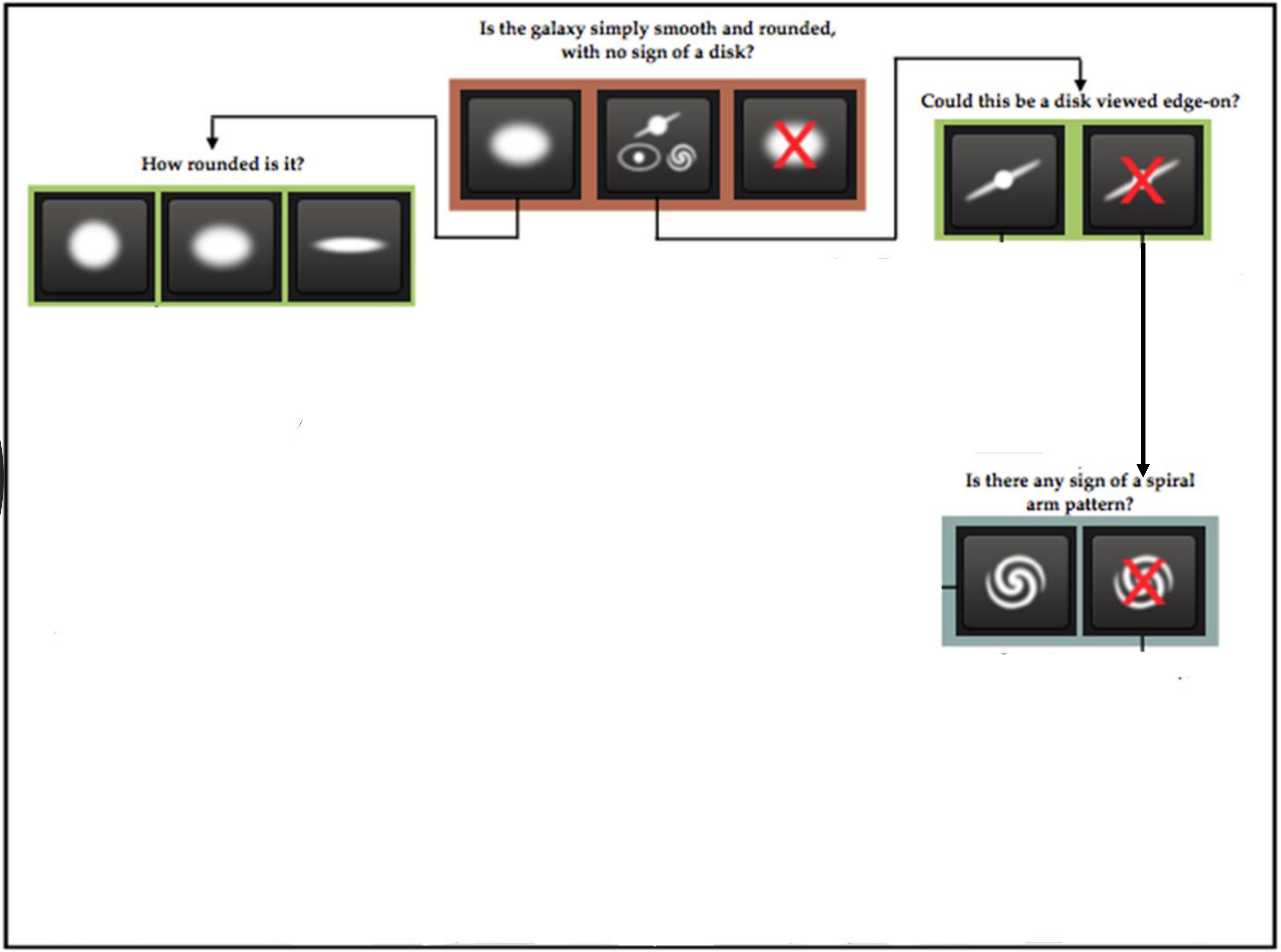
Q4. Is there a spiral pattern? *2 responses*

Q7. How round is the smooth galaxy? *3 responses*

The question(s)


Is the galaxy simply smooth and rounded, with no sign of a disk?
Could this be a disk viewed edge-on?
How rounded is it?
Is there a sign of a bar feature through the centre of the galaxy?
Does the galaxy have a bulge at its centre? If so, what shape?
Is there anything odd?
Is there any sign of a spiral arm pattern?
How tightly wound do the spiral arms appear?
Is the odd feature a ring, or is the galaxy disturbed or irregular?
How many spiral arms are there?
How prominent is the central bulge, compared to the rest of the galaxy?

The question(s)



Is the galaxy simply smooth and rounded, with no sign of a disk?

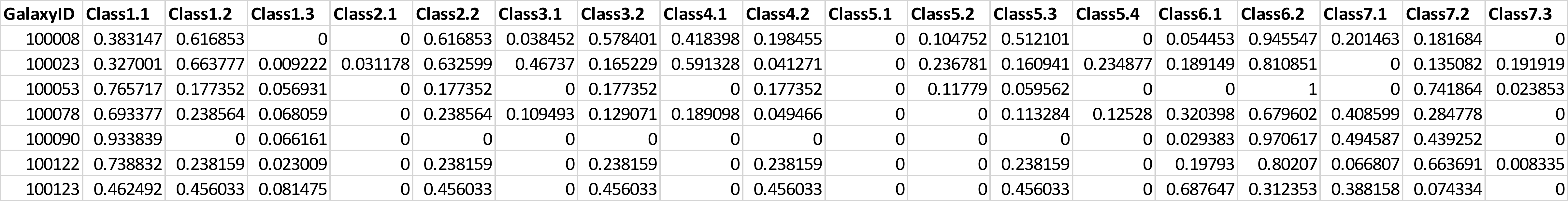
How rounded is it?

Could this be a disk viewed edge-on?

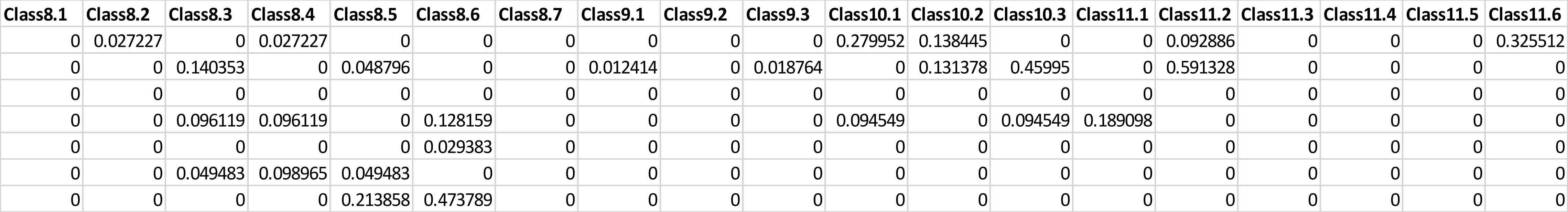
Is there any sign of a spiral arm pattern?

The data file

Each variable is the average percentage of volunteers’ responses to each question

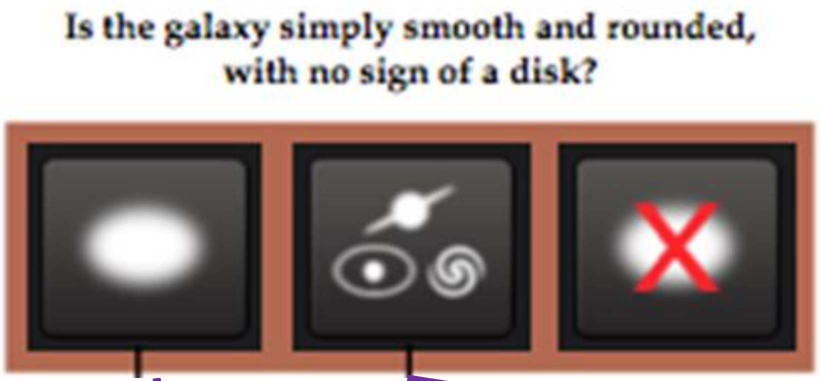


GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0

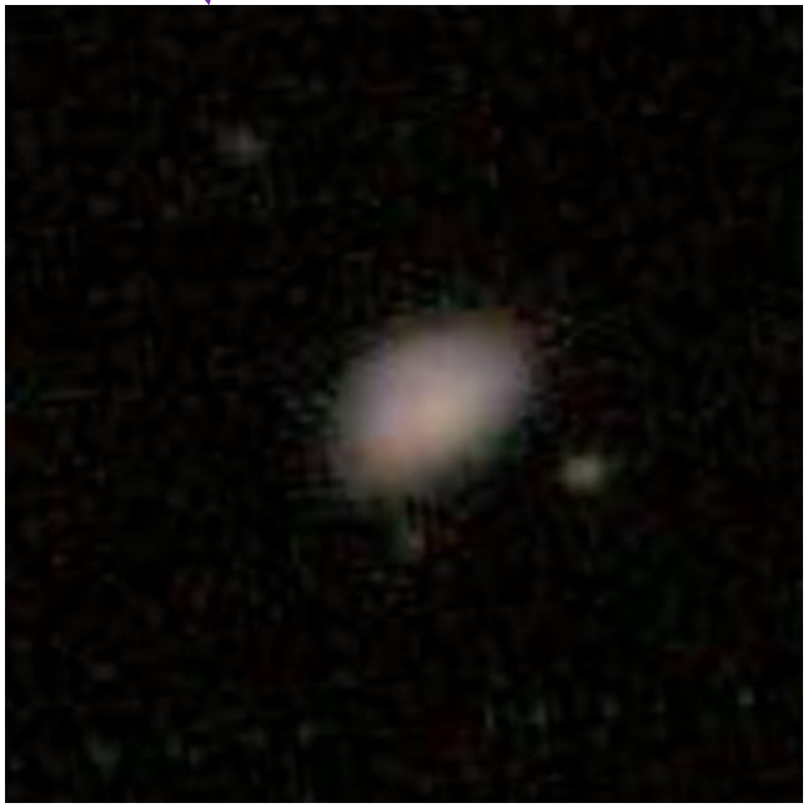


Class8.1	Class8.2	Class8.3	Class8.4	Class8.5	Class8.6	Class8.7	Class9.1	Class9.2	Class9.3	Class10.1	Class10.2	Class10.3	Class11.1	Class11.2	Class11.3	Class11.4	Class11.5	Class11.6
0	0.027227	0	0.027227	0	0	0	0	0	0	0.279952	0.138445	0	0	0.092886	0	0	0	0.325512
0	0	0.140353	0	0.048796	0	0	0.012414	0	0.018764	0	0.131378	0.45995	0	0.591328	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.096119	0.096119	0	0.128159	0	0	0	0	0.094549	0	0.094549	0.189098	0	0	0	0	0
0	0	0	0	0	0.029383	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.049483	0.098965	0.049483	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.213858	0.473789	0	0	0	0	0	0	0	0	0	0	0	0	0

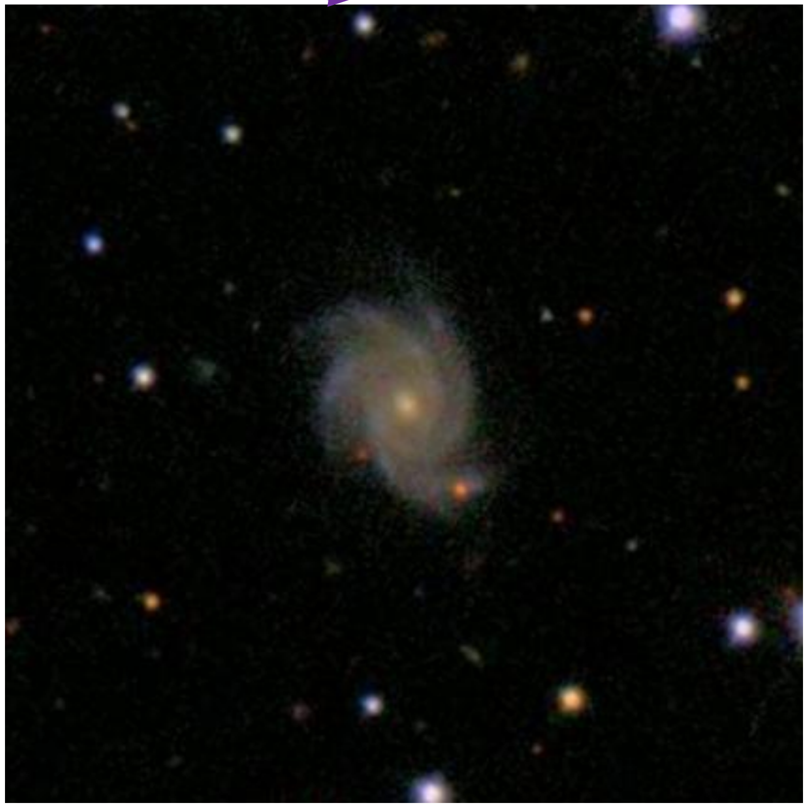
First Level



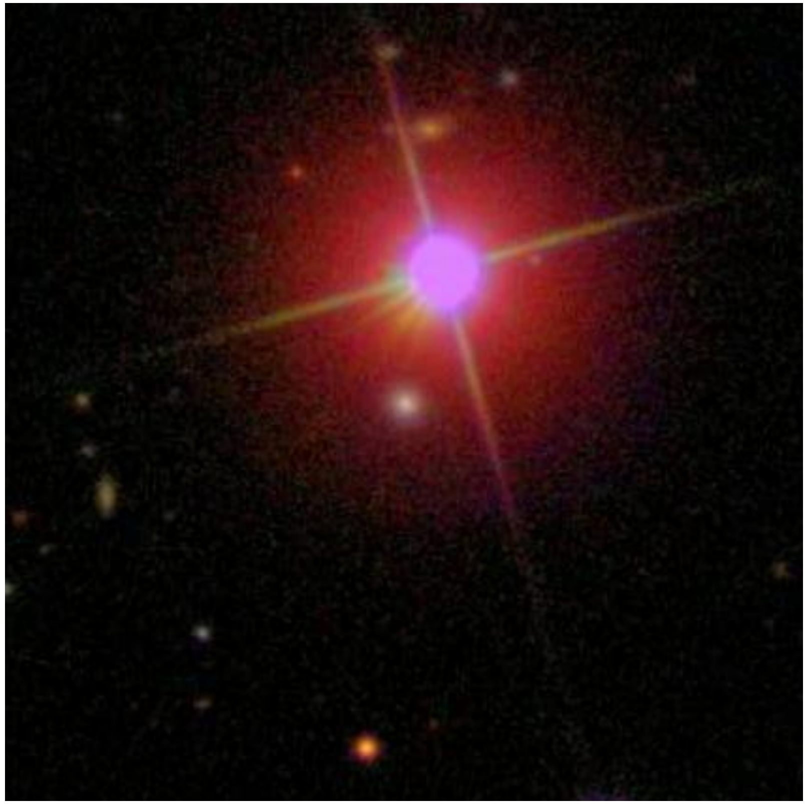
Is the galaxy simply smooth and rounded, with no sign of a disk?



Smooth/Rounded

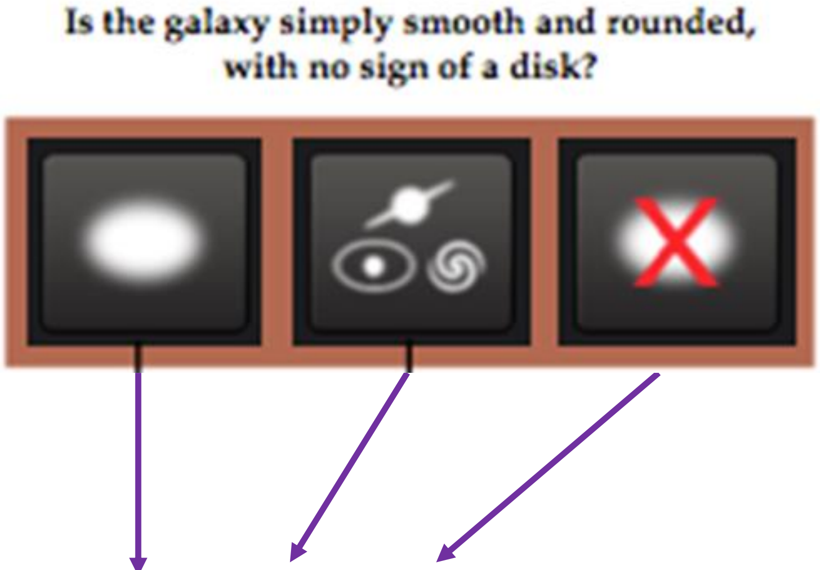


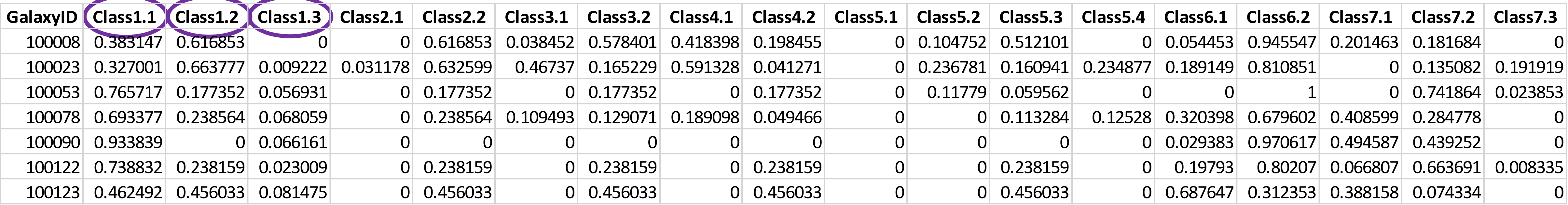
Features/Disk



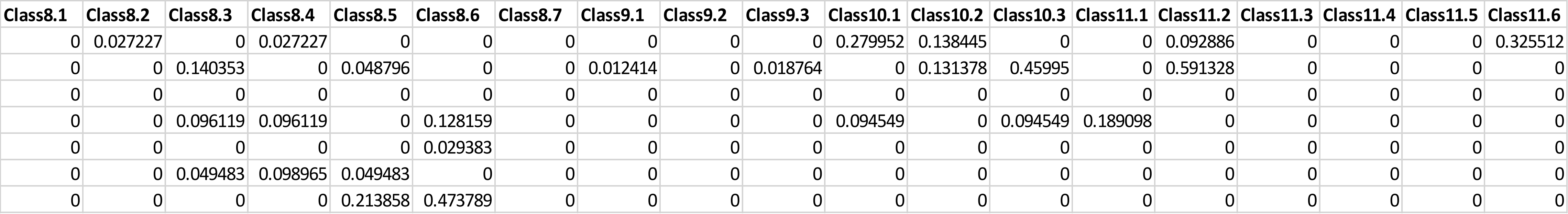
Not a Galaxy

First Level


Is the galaxy simply smooth and rounded, with no sign of a disk?

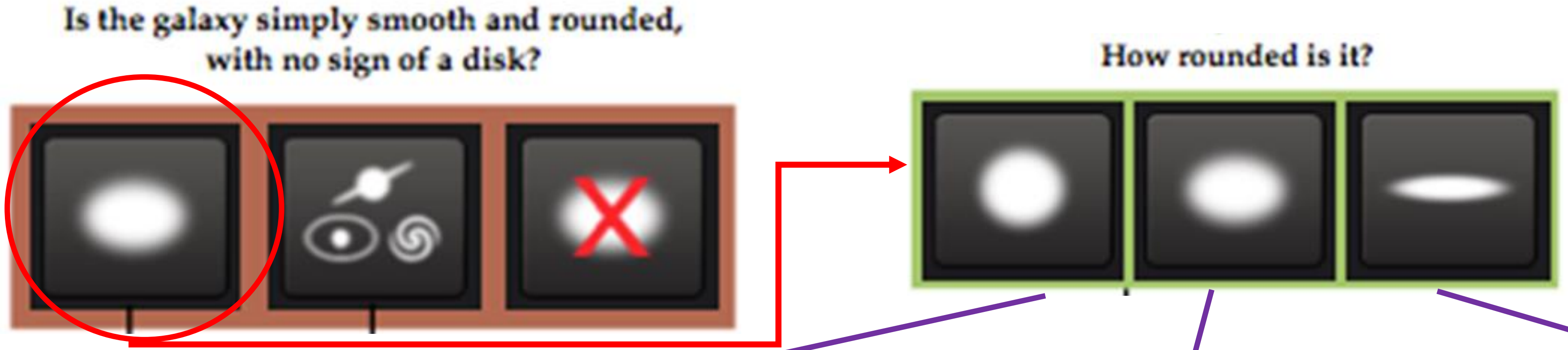


GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0



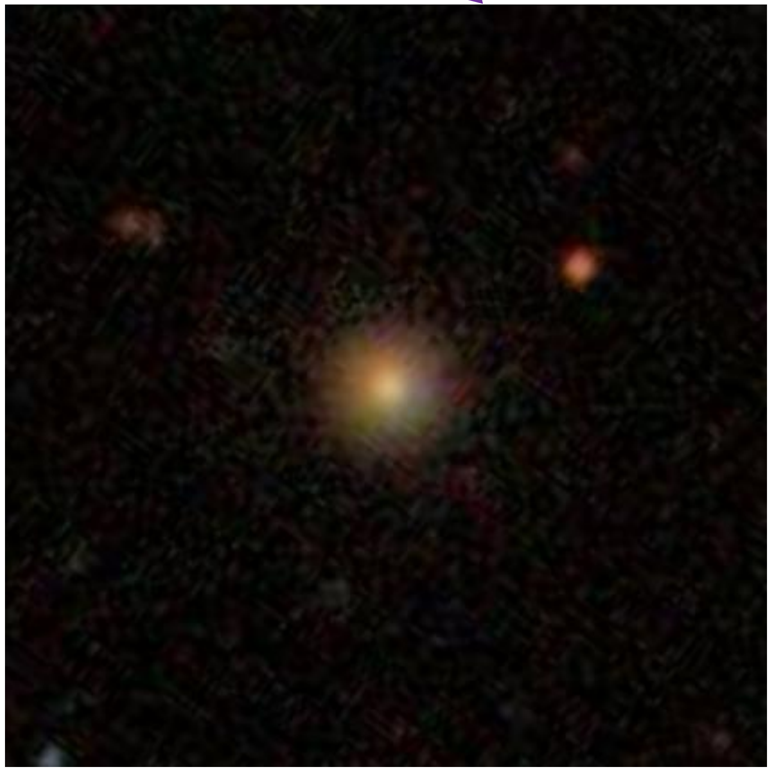
Class8.1	Class8.2	Class8.3	Class8.4	Class8.5	Class8.6	Class8.7	Class9.1	Class9.2	Class9.3	Class10.1	Class10.2	Class10.3	Class11.1	Class11.2	Class11.3	Class11.4	Class11.5	Class11.6
0	0.027227	0	0.027227	0	0	0	0	0	0	0.279952	0.138445	0	0	0.092886	0	0	0	0.325512
0	0	0.140353	0	0.048796	0	0	0.012414	0	0.018764	0	0.131378	0.45995	0	0.591328	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.096119	0.096119	0	0.128159	0	0	0	0	0.094549	0	0.094549	0.189098	0	0	0	0	0
0	0	0	0	0	0.029383	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.049483	0.098965	0.049483	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.213858	0.473789	0	0	0	0	0	0	0	0	0	0	0	0	0

Second Level



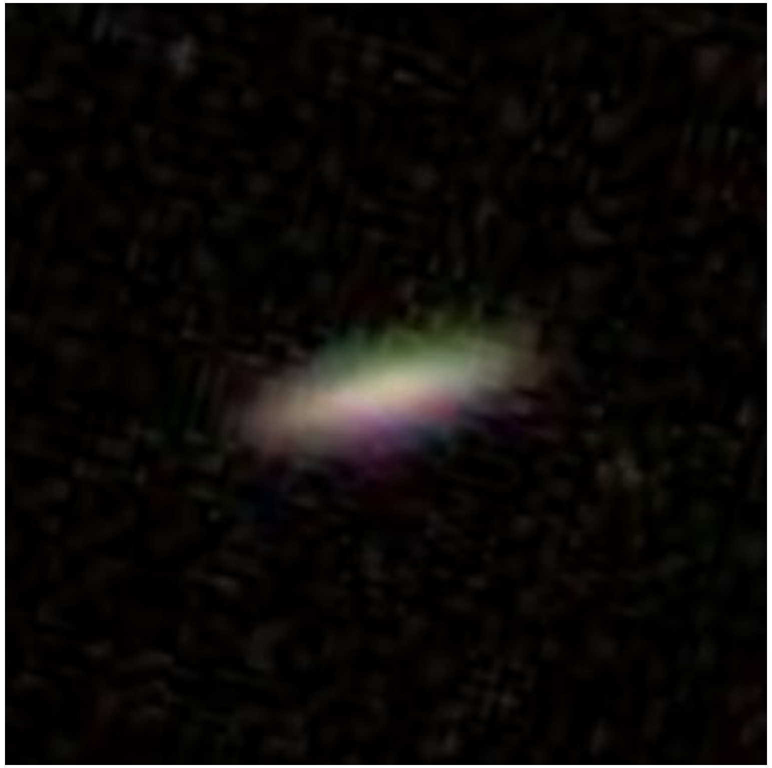
Is the galaxy simply smooth and rounded, with no sign of a disk?

How rounded is it?



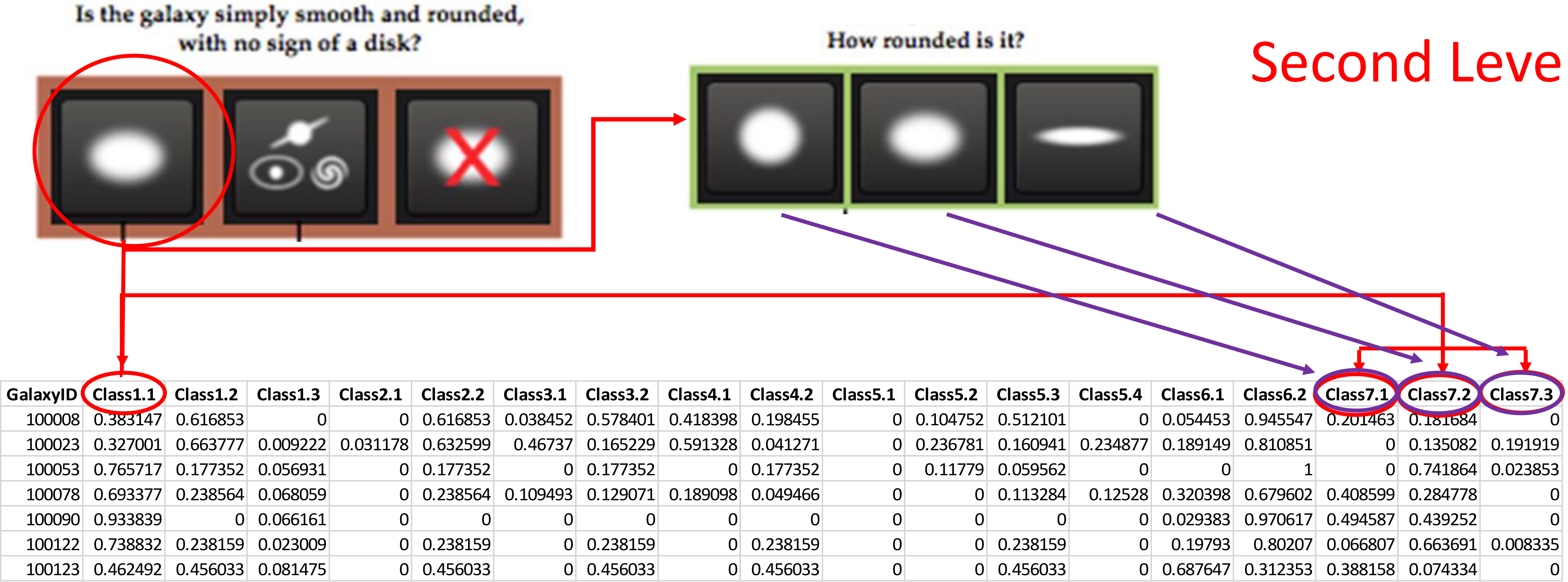
Completely Round

Oval



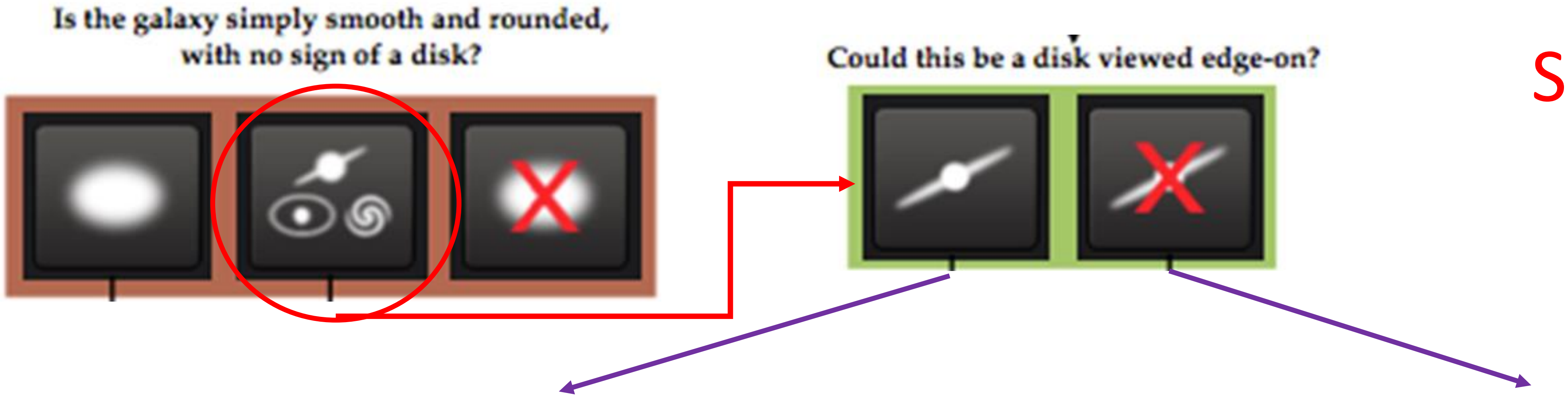
Cigar-shaped

Second Level


Is the galaxy simply smooth and rounded, with no sign of a disk?
How rounded is it?

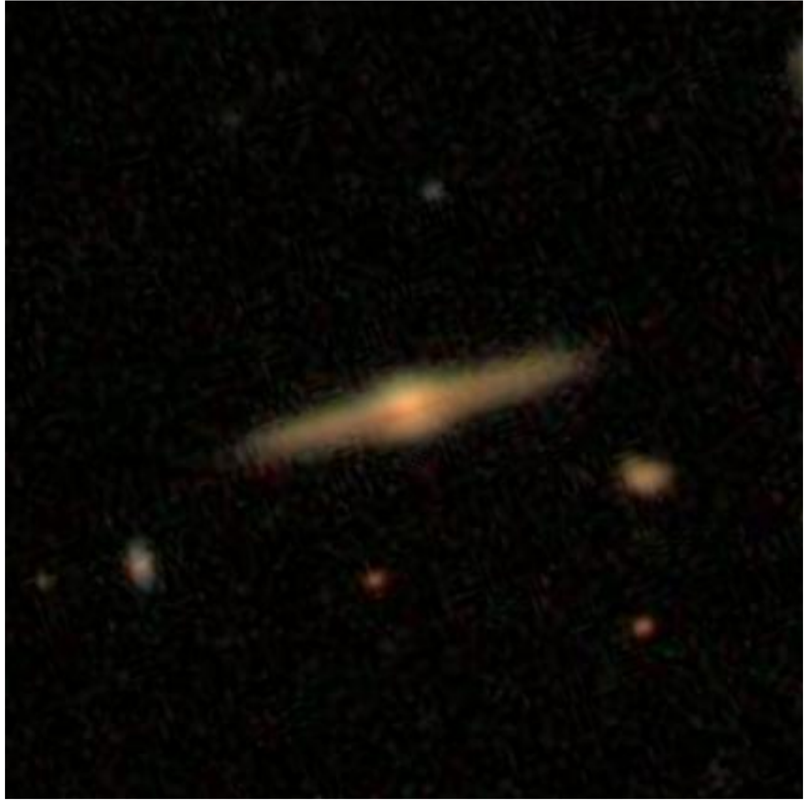
GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0

Second Level

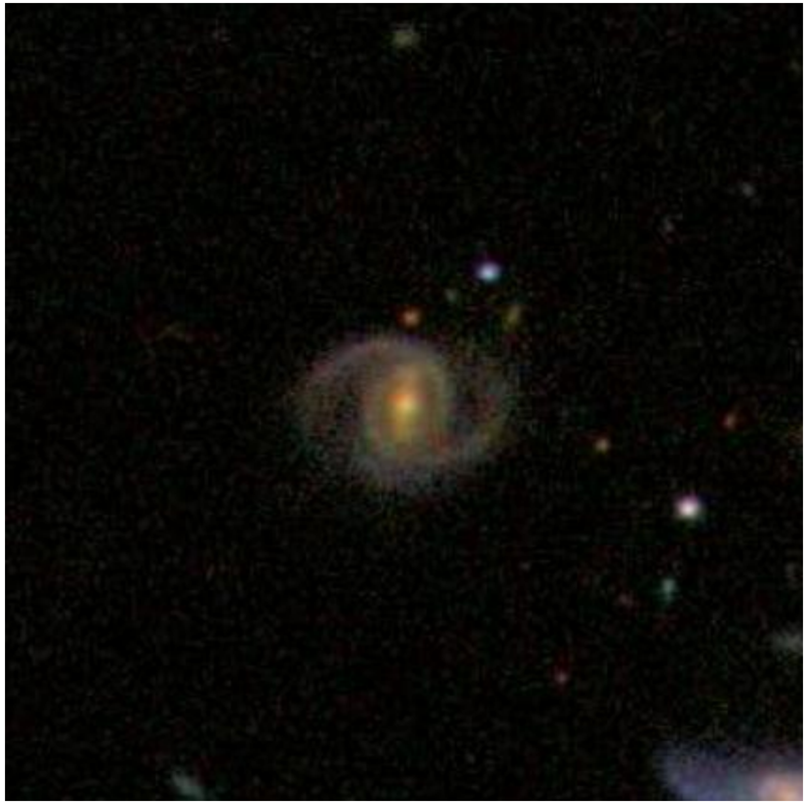


Is the galaxy simply smooth and rounded, with no sign of a disk?

Could this be a disk viewed edge-on?

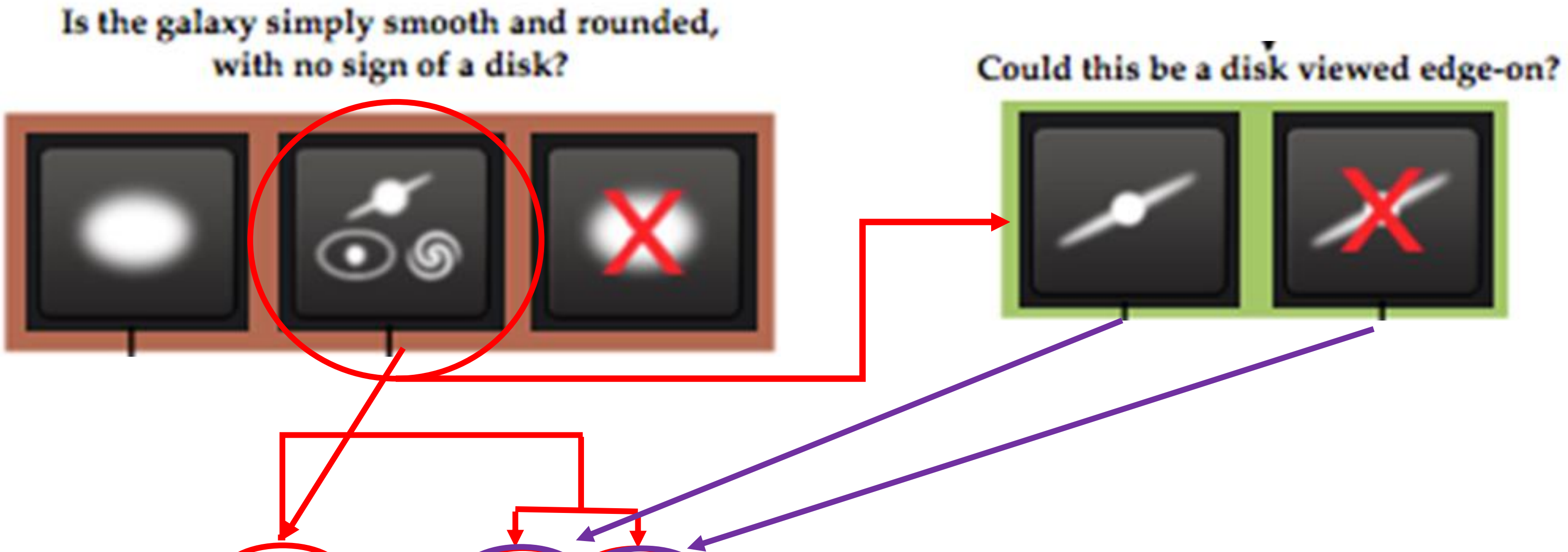


Edge-on Disk



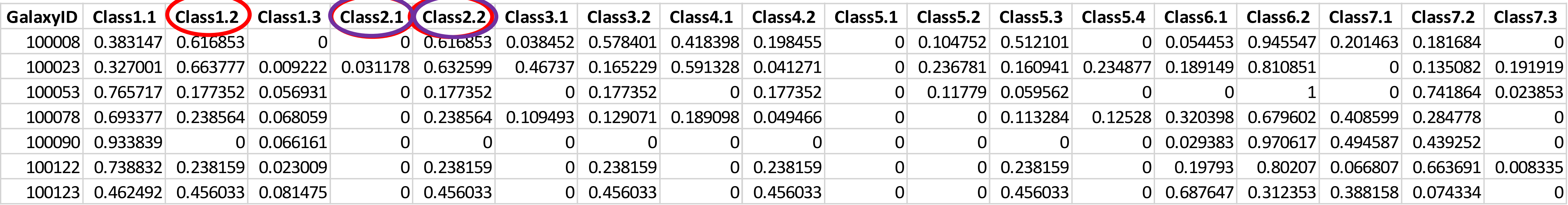
Not Edge-on

Second Level



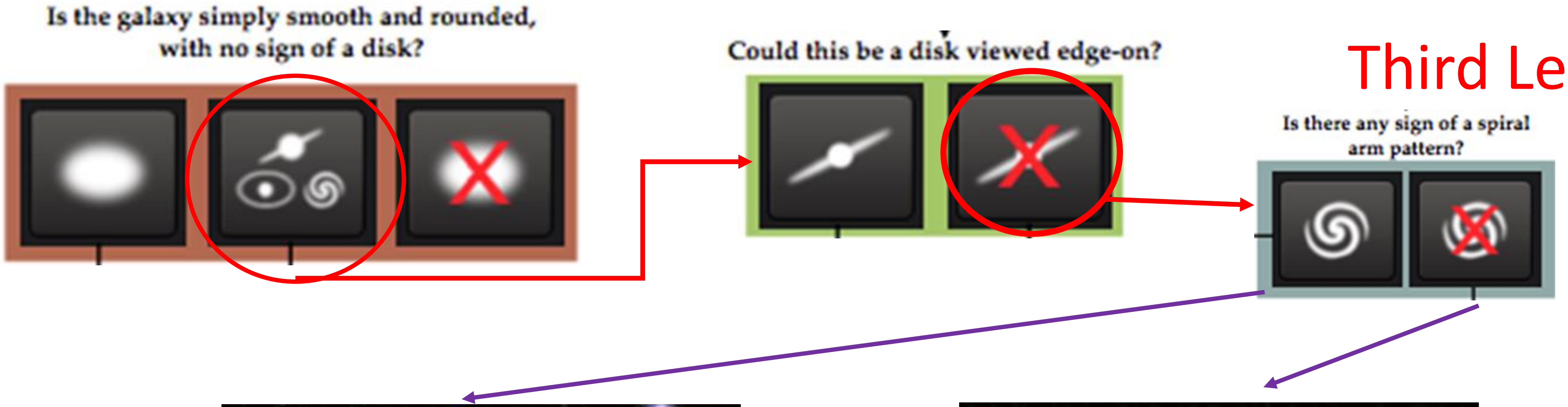
Is the galaxy simply smooth and rounded, with no sign of a disk?

Could this be a disk viewed edge-on?



GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0

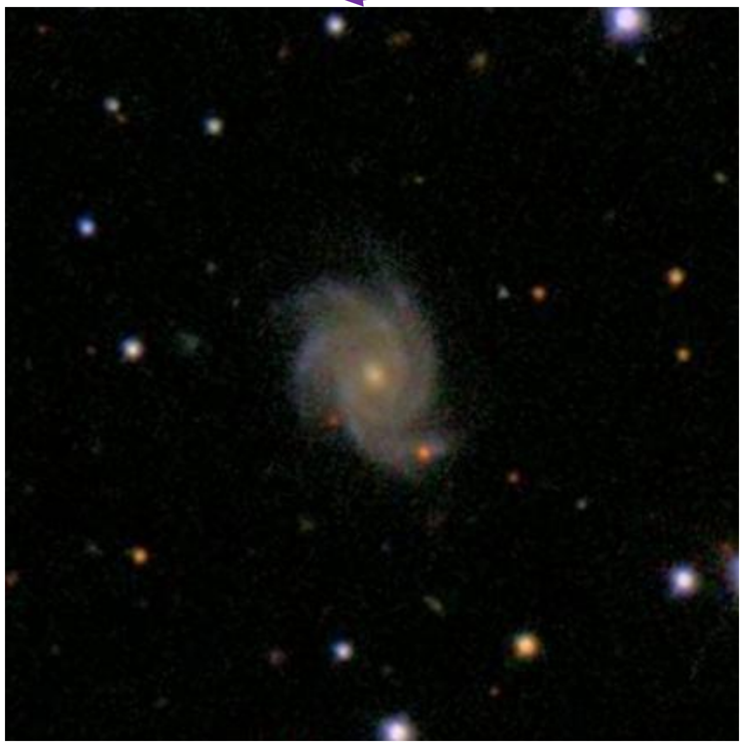
Third Level



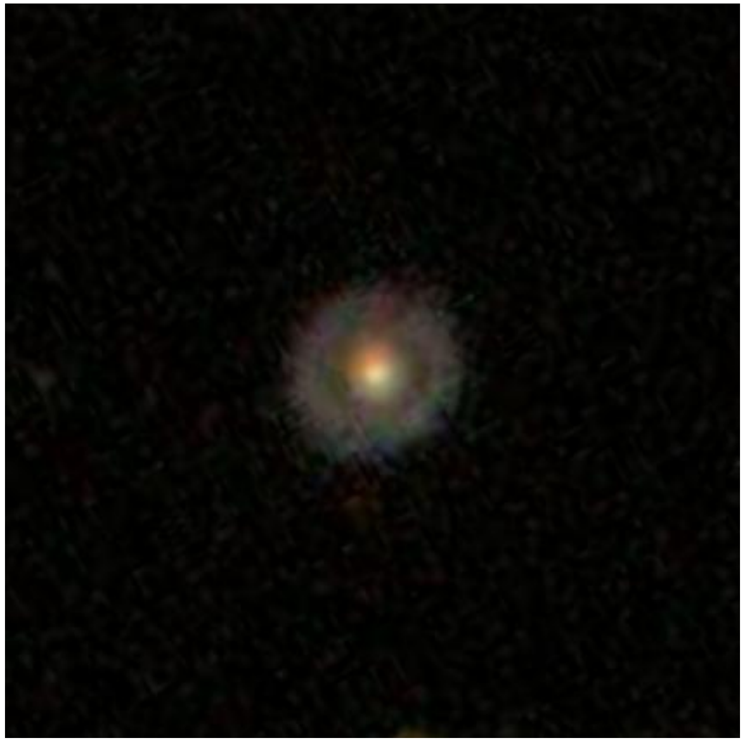
Is the galaxy simply smooth and rounded, with no sign of a disk?

Could this be a disk viewed edge-on?

Is there any sign of a spiral arm pattern?

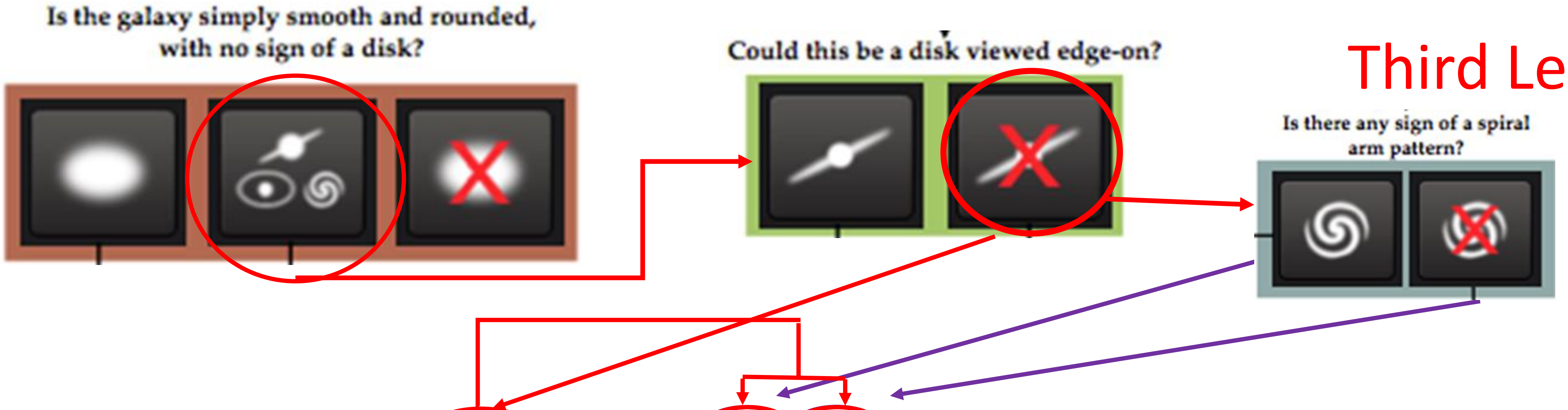


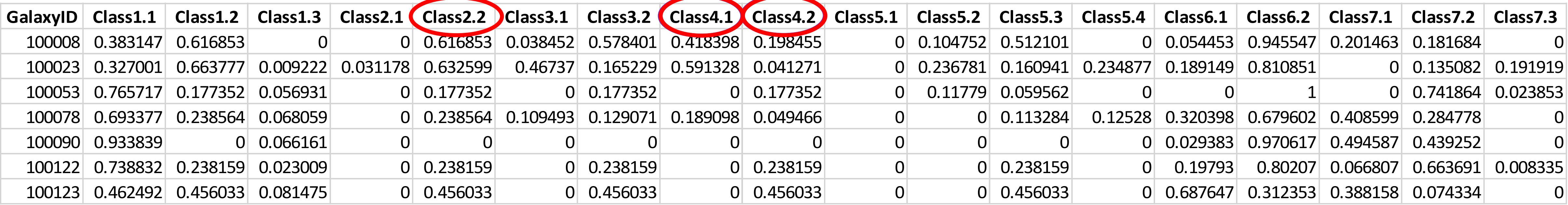
Spiral pattern



Non-spiral pattern

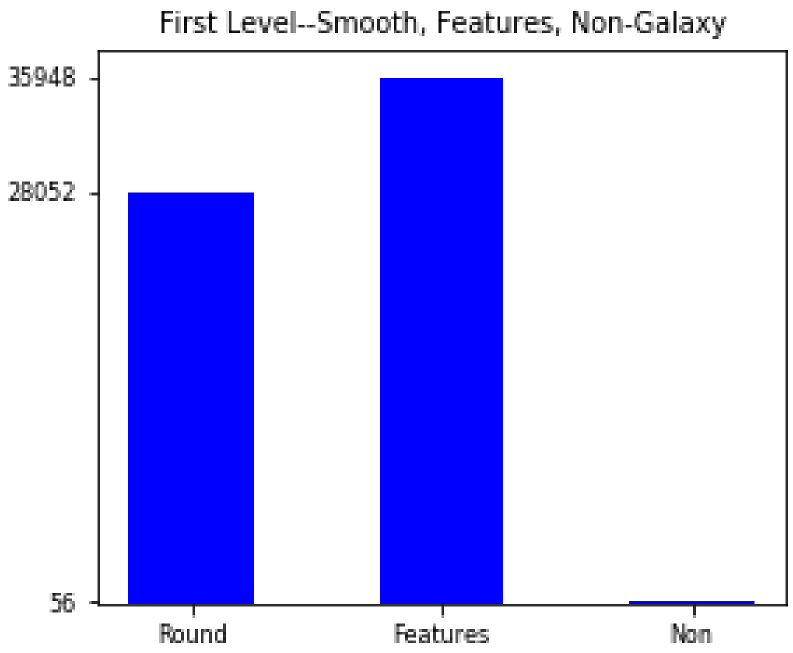
Third Level

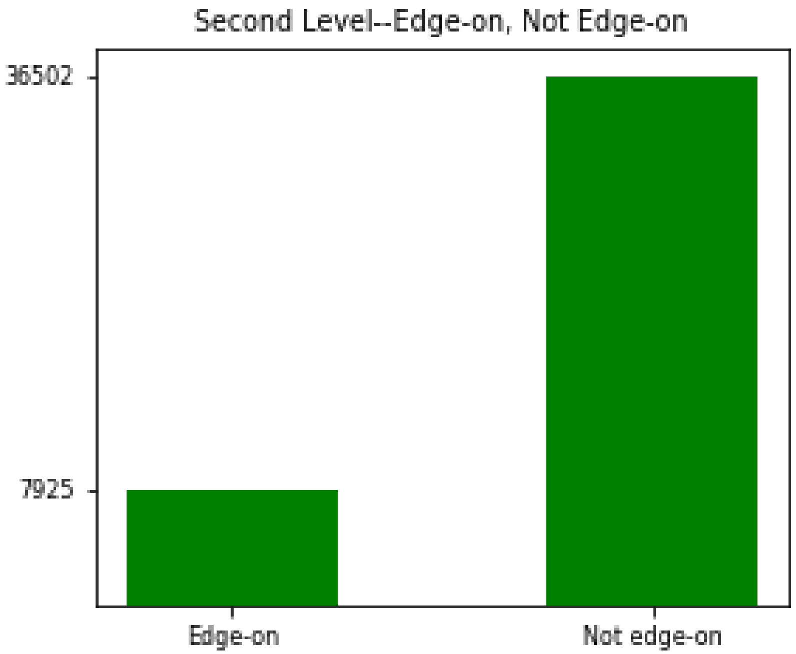

Is the galaxy simply smooth and rounded, with no sign of a disk?
Could this be a disk viewed edge-on?
Is there any sign of a spiral arm pattern?

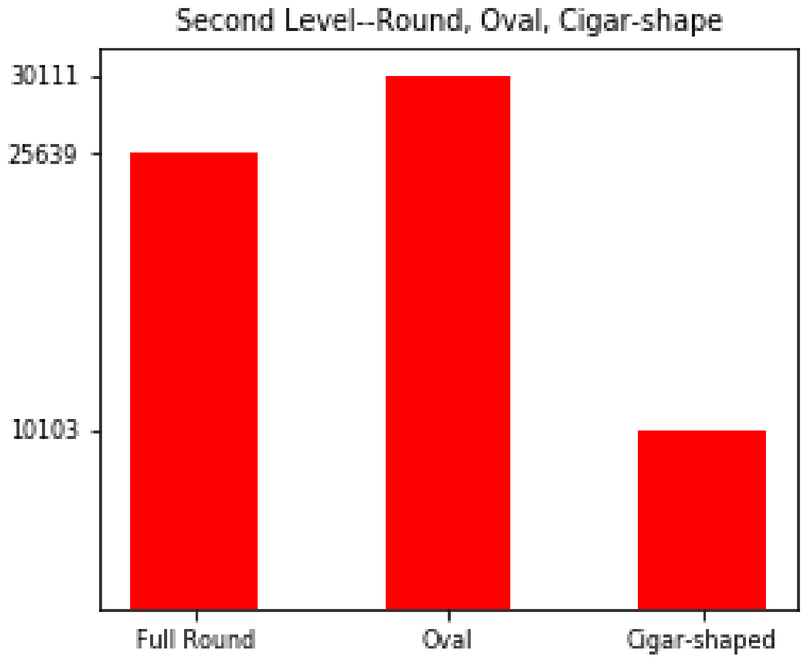


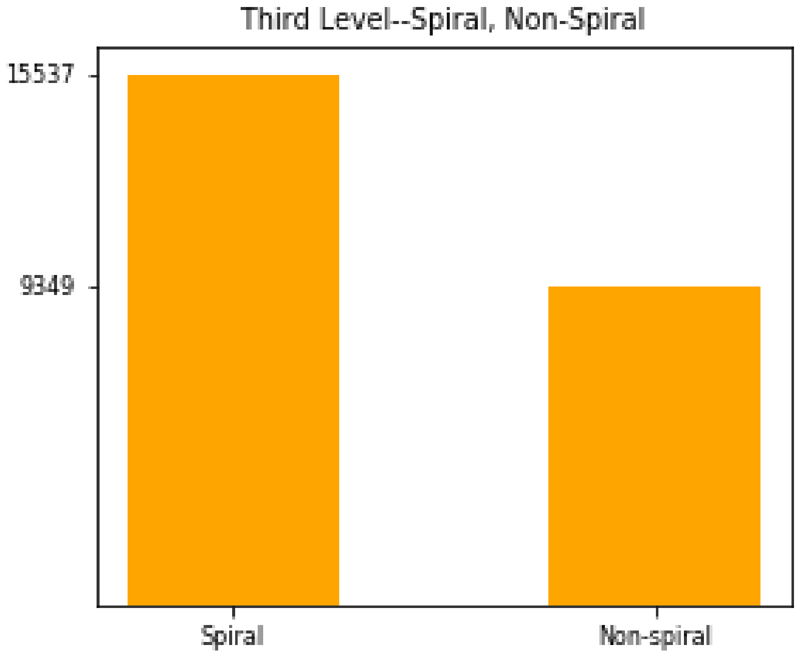
GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0

Image Distribution


First Level--Smooth, Features, Non-Galaxy

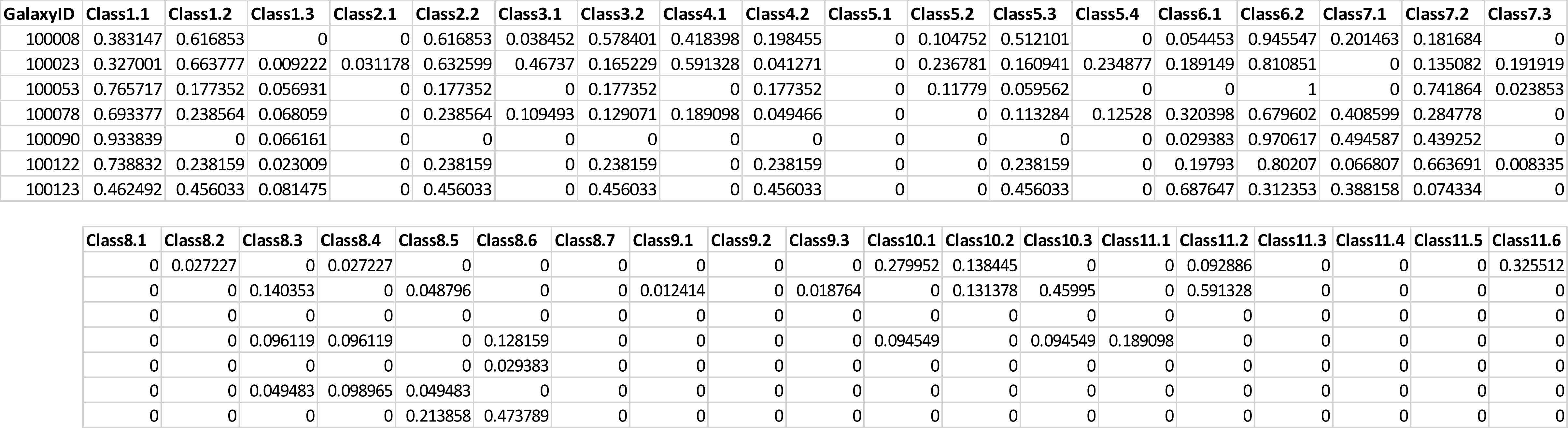

Second Level--Edge-on, Not Edge-on


Second Level--Round, Oval, Cigar-shape


Third Level--Spiral, Non-Spiral

The data file

Each variable is the average percentage of volunteers’ responses to each question

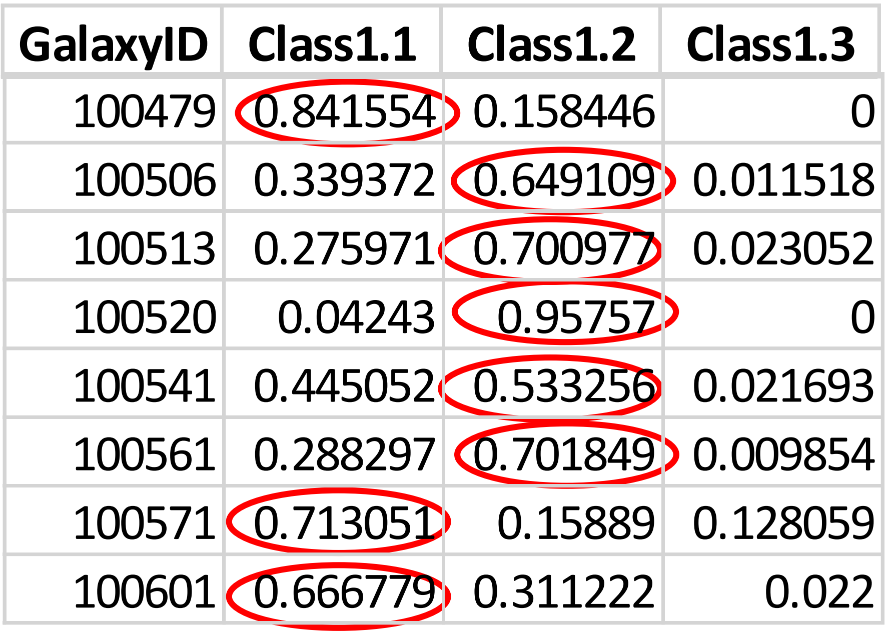


GalaxyID	Class1.1	Class1.2	Class1.3	Class2.1	Class2.2	Class3.1	Class3.2	Class4.1	Class4.2	Class5.1	Class5.2	Class5.3	Class5.4	Class6.1	Class6.2	Class7.1	Class7.2	Class7.3	Class8.1	Class8.2	Class8.3	Class8.4	Class8.5	Class8.6	Class8.7	Class9.1	Class9.2	Class9.3	Class10.1	Class10.2	Class10.3	Class11.1	Class11.2	Class11.3	Class11.4	Class11.5	Class11.6
100008	0.383147	0.616853	0	0	0.616853	0.038452	0.578401	0.418398	0.198455	0	0.104752	0.512101	0	0.054453	0.945547	0.201463	0.181684	0	0	0.027227	0	0.027227	0	0	0	0	0	0	0.279952	0.138445	0	0	0.092886	0	0	0	0.325512
100023	0.327001	0.663777	0.009222	0.031178	0.632599	0.46737	0.165229	0.591328	0.041271	0	0.236781	0.160941	0.234877	0.189149	0.810851	0	0.135082	0.191919	0	0	0.140353	0	0.048796	0	0	0.012414	0	0.018764	0	0.131378	0.45995	0	0.591328	0	0	0	0
100053	0.765717	0.177352	0.056931	0	0.177352	0	0.177352	0	0.177352	0	0.11779	0.059562	0	0	1	0	0.741864	0.023853	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100078	0.693377	0.238564	0.068059	0	0.238564	0.109493	0.129071	0.189098	0.049466	0	0	0.113284	0.12528	0.320398	0.679602	0.408599	0.284778	0	0	0	0.096119	0.096119	0	0.128159	0	0	0	0	0.094549	0	0.094549	0.189098	0	0	0	0	0
100090	0.933839	0	0.066161	0	0	0	0	0	0	0	0	0	0	0.029383	0.970617	0.494587	0.439252	0	0	0	0	0	0	0.029383	0	0	0	0	0	0	0	0	0	0	0	0	0
100122	0.738832	0.238159	0.023009	0	0.238159	0	0.238159	0	0.238159	0	0	0.238159	0	0.19793	0.80207	0.066807	0.663691	0.008335	0	0	0.049483	0.098965	0.049483	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100123	0.462492	0.456033	0.081475	0	0.456033	0	0.456033	0	0.456033	0	0	0.456033	0	0.687647	0.312353	0.388158	0.074334	0	0	0	0	0	0.213858	0.473789	0	0	0	0	0	0	0	0	0	0	0	0	0

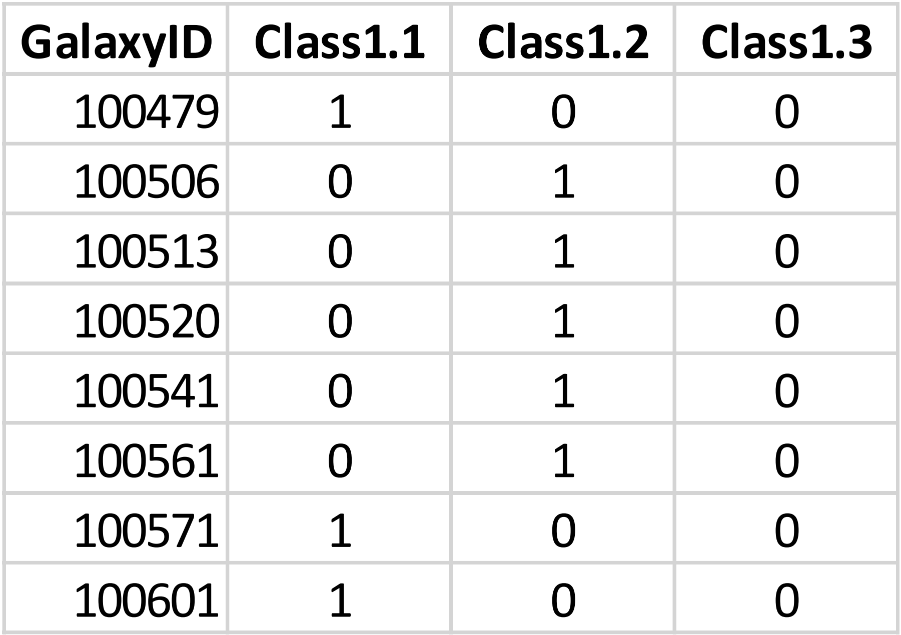
- The original Kaggle competition focused on average MSE for each response
- My project is changing these to categoricals and predicting the category

Data wrangling

Convert results of classes into categoricals – maximum value



	Round	Features	Non-galaxy
GalaxyID	**Class1.1**	**Class1.2**	**Class1.3**
100479	0.841554	0.158446	0
100506	0.339372	0.649109	0.011518
100513	0.275971	0.700977	0.023052
100520	0.04243	0.95757	0
100541	0.445052	0.533256	0.021693
100561	0.288297	0.701849	0.009854
100571	0.713051	0.15889	0.128059
100601	0.666779	0.311222	0.022



	Round	Features	Non-galaxy
GalaxyID	**Class1.1**	**Class1.2**	**Class1.3**
100479	1	0	0
100506	0	1	0
100513	0	1	0
100520	0	1	0
100541	0	1	0
100561	0	1	0
100571	1	0	0
100601	1	0	0

More Tests

Multiple tests based on response percentages

- Maximum value per category (entire dataset)
- At least 50% average for one category
- 60%, 70%, 80%, 90%

Data wrangling

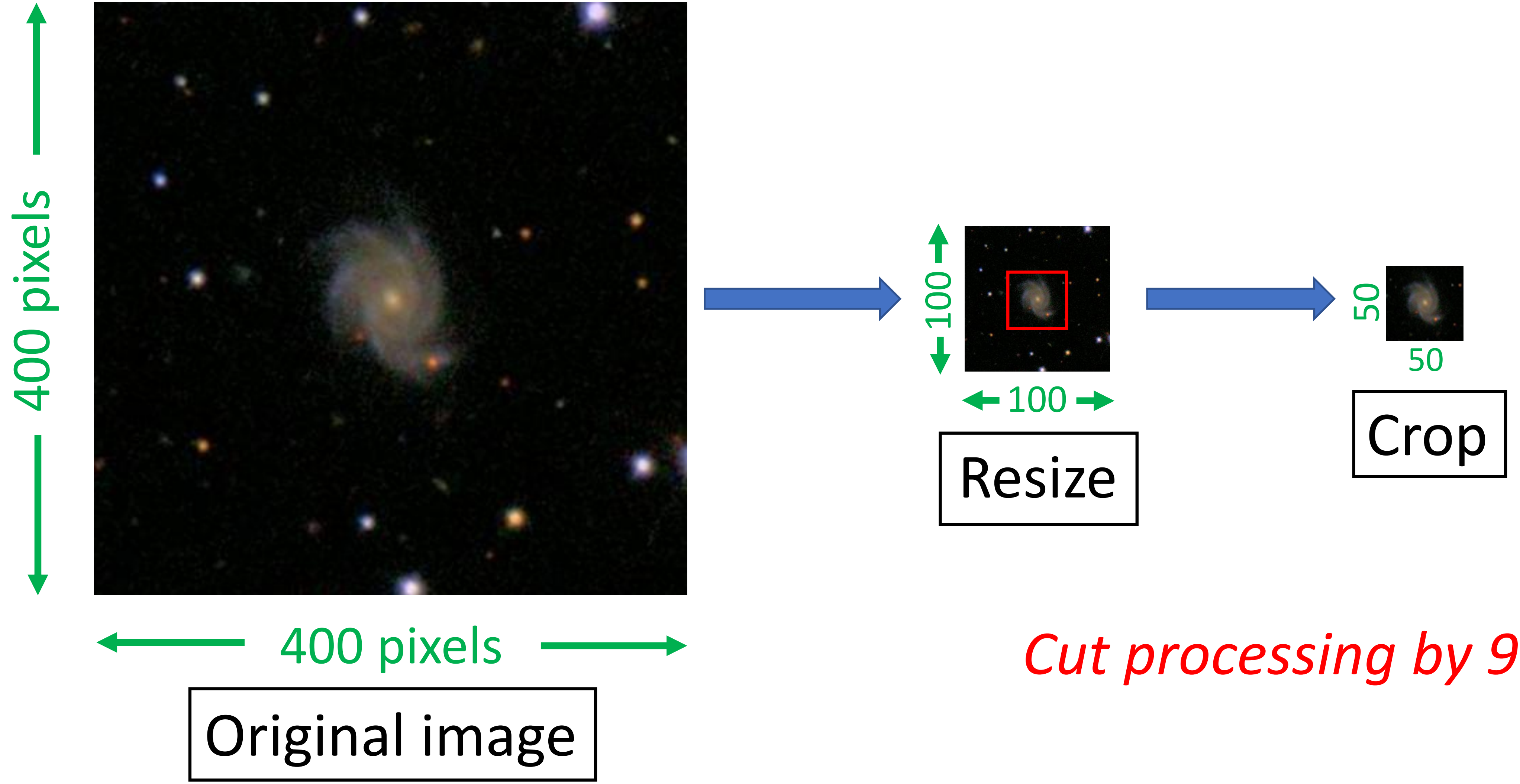
Convert results of classes into categoricals – 70% minimum result

- Drop all rows where the highest value < 70%

	Round	Features	Non-galaxy
GalaxyID	**Class1.1**	**Class1.2**	**Class1.3**
100479	0.841554	0.158446	0
~~100506~~	~~0.339372~~	~~0.649109~~	~~0.011518~~
100513	0.275971	0.700977	0.023052
100520	0.04243	0.95757	0
~~100541~~	~~0.445052~~	~~0.533256~~	~~0.021693~~
100561	0.288297	0.701849	0.009854
100571	0.713051	0.15889	0.128059
~~100601~~	~~0.666779~~	~~0.311222~~	~~0.022~~

GalaxyID	Class1.1	Class1.2	Class1.3
100479	1	0	0
100513	0	1	0
100520	0	1	0
100561	0	1	0
100571	1	0	0

Image wrangling with OpenCV



400 pixels

400 pixels

Original image

100

100

Resize

50

50

Crop

Cut processing by 98.5%

Modeling approach

- Modeling data
  - Training: 40K images
  - Validation: 10K images
  - Testing: 10K images
- Data conversion
  - OpenCV to convert the images to arrays
  - Pickled the image arrays once captured
    - 0.3 seconds vs. 6 minutes
    - No need to read and process every image every time
- Convert all values in solutions file to categoricals
- Compared 2 neural network models
  - Multi-Layer Perceptron (MLP)
  - Convolutional
- Both performed well in initial tests
  - Convolutional, although slower, on average was about 5% better
  - Go with convolutional and tweak

Modeling

- Convolutional parameters
  - Optimizers: RMSProp, SGD, Adam, Adagrad, Adadelta, Adam, Adamax
  - Loss function: Categorical Crossentropy (standard)
  - Number of Epochs
    - Trained between 5 and 60 epochs for analysis
    - Used EarlyStopping technique (stop training after 4 epochs of no improvement)
    - 20 epochs on average before model stopped improving
  - Neurons/layer
    - Start with 32 or 64, and double in size
    - Little difference—in fact, starting with 64 sometimes gave worse results and took MUCH longer to run
  - Number of layers
    - Simple 3-layer
    - VGG-16-like technique (multiple convolutional layers and maxpooling layers)
  - Batch-size
    - 64 or 128

Modeling

Final model

- Optimizer: Adamax
- Number of Epochs: Early Stopping (usually around 20)
- Batch size: 64
- VGG-16 like technique
  - 2 layers of 32, maxpool, drop 0.2
  - 3 layers of 64, maxpool, drop 0.2
  - 4 layers of 128, maxpool, drop 0.2
  - Flatten, then dense layer of 256, drop 0.5
  - Dense layer of number of categories predicted

Results

Maximum response per category

- First level (smooth/features/non-galaxy)
  - Train images: 40,000   Val: 10,000   Test: 10,000
  - Validation: 87.0 %
  - Test: 86.3%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 40,000   Val: 10,000   Test: 10,000
  - Validation: 76.1%
  - Test: 76.0%
- Third level (spiral/non-spiral)
  - Train images: 40,000   Val: 10,000   Test: 10,000
  - Validation: 62.6%
  - Test: 62.6%

Results

At least 50% response per category

- First level (smooth/features/non-galaxy)
  - Train images: 39,010 Val: 9,719 Test: 9,744
  - Validation: 86.2%
  - Test: 85.9%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 38,970 Val: 9,704 Test: 9,735
  - Validation: 79.4%
  - Test: 79.4%
- Third level (spiral/non-spiral)
  - Train images: 27,348 Val: 6,897 Test: 6,869
  - Validation: 72.6%
  - Test: 72.0%

Results

At least 60% response per category

- First level (smooth/features/non-galaxy)
  - Train images: 31,485 Val: 7,881 Test: 7,884
  - Validation: 91.2%
  - Test: 91.3%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 28,940 Val: 7,281 Test: 7,241
  - Validation: 87.6%
  - Test: 87.3%
- Third level (spiral/non-spiral)
  - Train images: 27,348 Val: 6,897 Test: 6,869
  - Validation: 78.5%
  - Test: 77.8%

Results

At least 70% response per category

- First level (smooth/features/non-galaxy)
  - Train images: 23,882 Val: 5,950 Test: 5,983
  - Validation: 96%
  - Test: 95.8%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 20,173 Val: 5,028 Test: 5,026
  - Validation: 93.4%
  - Test: 94%
- Third level (spiral/non-spiral)
  - Train images: 18,709 Val: 4,686 Test: 4,661
  - Validation: 82.3%
  - Test: 83.3%

Results

At least 80% response per category

- First level (smooth/features/non-galaxy)
  - Train images: 15,179   Val: 3,980   Test: 3,927
  - Validation: 97.5%
  - Test: 97.7%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 12,332   Val: 3,085   Test: 3,037
  - Validation: 96%
  - Test: 96.3%
- Third level (spiral/non-spiral)
  - Train images: 10,044   Val: 2,516   Test: 2,471
  - Validation: 90.7%
  - Test: 92.1%

Results

At least 90% response per category

- First level (smooth/features/non-galaxy)
  - Train images: 7,323 Val: 1,828 Test: 1,845
  - Validation: 98.9%
  - Test: 98.9%
- Second level (1. edge-on/not-edge-on; 2. how round)
  - Train images: 5,526 Val: 1,365 Test: 1,360
  - Validation: 98%
  - Test: 98.5%
- Third level (spiral/non-spiral)
  - Train images: 3,966 Val: 941 Test: 989
  - Validation: 96.2%
  - Test: 97.2%

Conclusion

- Convnet did extremely well for subjective categories
- This model can be used to provide high accuracy predictions of galaxy types and features
- Many other things/permutations to explore
  - More images?
  - More categories
  - Play with percentages

Challenges

- So many parameters, so little time
  - Need to try a few out with subsets of the data
  - Balance accuracy with computation time
- Large data set
- Google Cloud Platform
  - Signed up for free GCP account ($300 credit)
  - Can use many vCPU’s and vGPU’s
  - Burned a lot of time wrestling with issues
- Intense processing crashed Jupyter many times

Thanks for watching!