

A method for utilizing automated machine learning for histopathological classification of testis based on Johnsen scores

Yurika Ito¹, Mami Unagami¹, Fumito Yamabe¹, Yozo Mitsui¹, Koichi Nakajima¹, Koichi Nagao¹, Hideyuki Kobayashi^{1,*}

¹ Department of Urology, Toho University School of Medicine, Tokyo, Japan

*To whom correspondence should be addressed: Department of Urology, Toho University School of Medicine, Tokyo, Japan, 6-11-1, Omori-Nishi, Ota-ku, Tokyo 143-8541, Japan

TEL +81-3-3762-4151 FAX +81-3-3768-8817

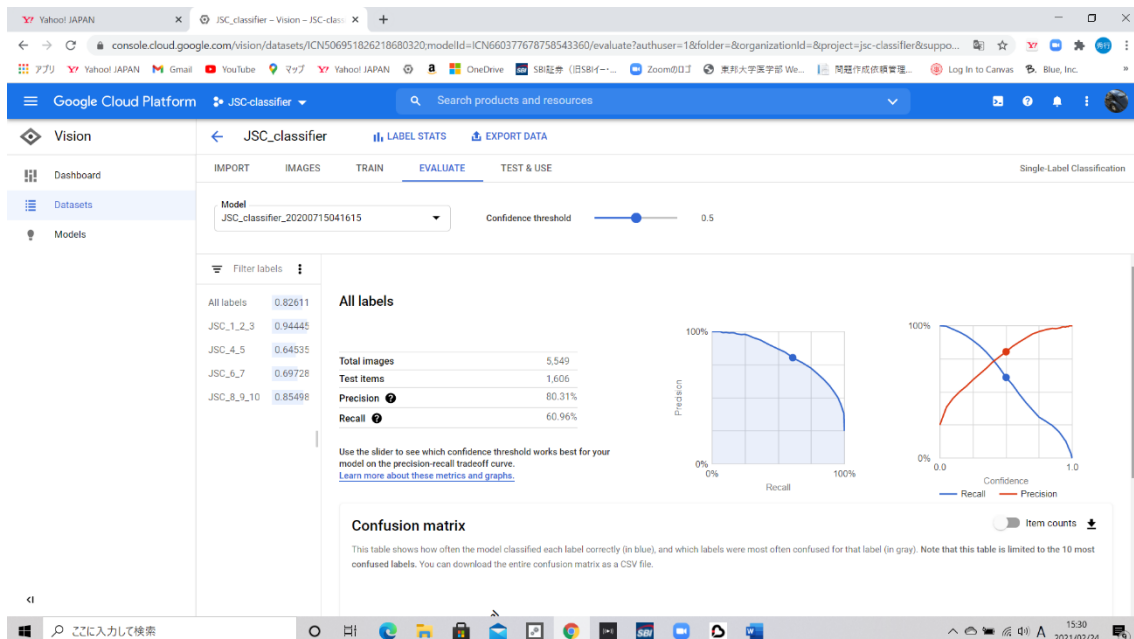
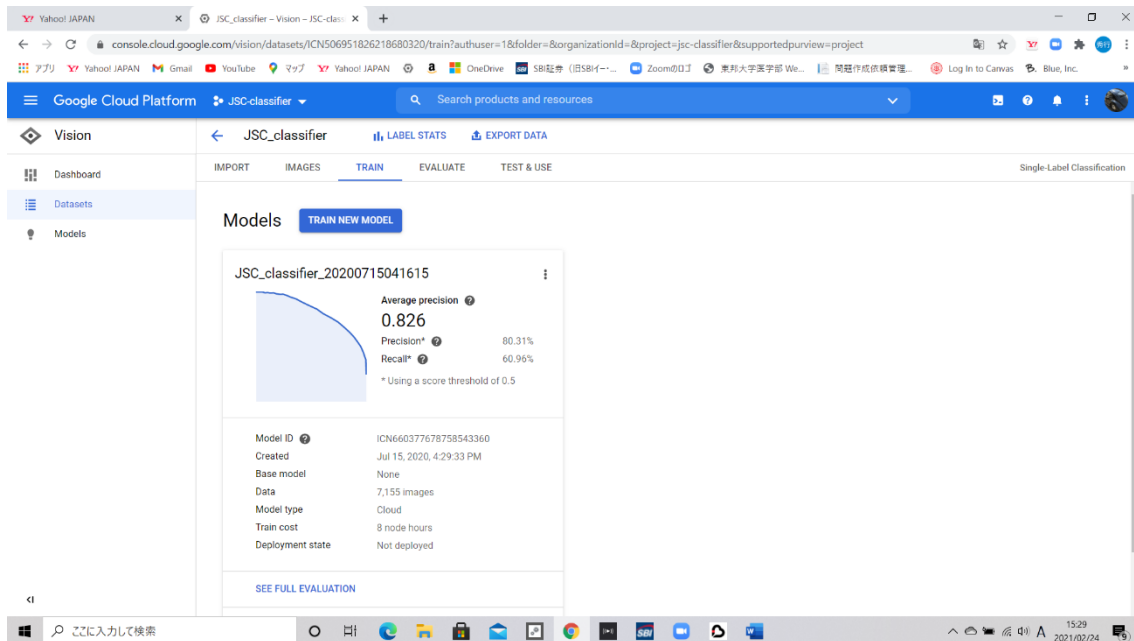
E-mail: hideyukk@med.toho-u.ac.jp

Supplementary data

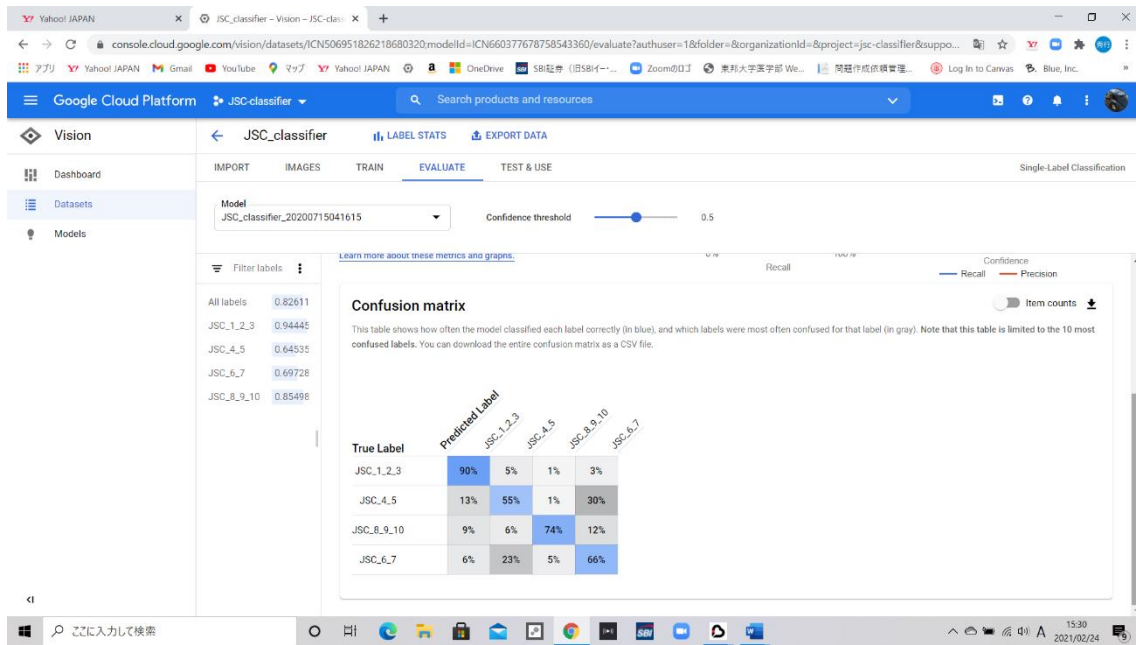
All statistical analyses for JSC_classifier: an image dataset of 7155 magnification images (X400) and JSC_classifier_Part2: an image dataset of 9822 expansion images (5.0 X 5.0 cm)

JSC_classifier: an image dataset of 7155 magnification images (X400)

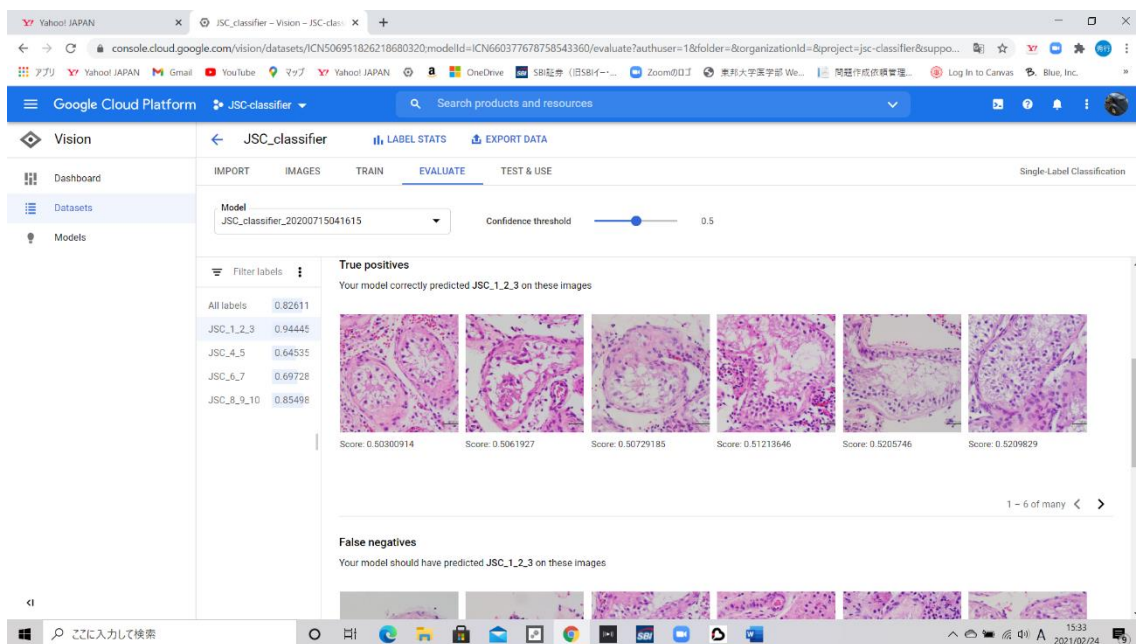
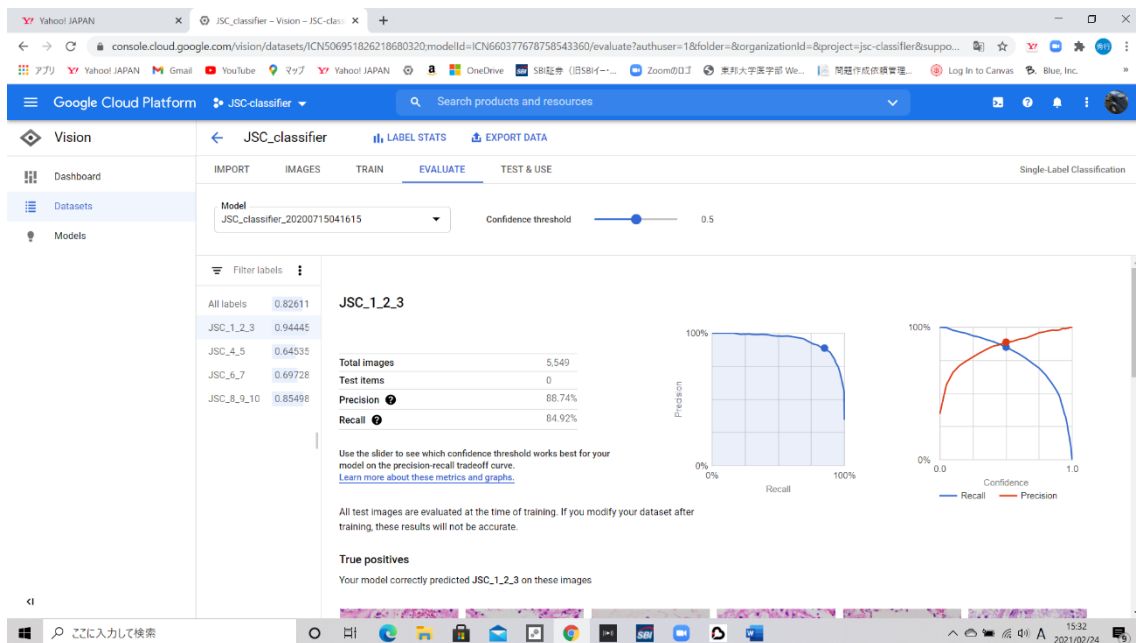
The average precision (positive predictive value) of the algorithm was 82.6%, precision was 80.31%, and recall was 60.96% for all labels.



Confusion matrix



The precision was 88.74% and recall was 84.92% for a Johnsen score of 1-3. True positive, false negative, and false positive images are indicated.

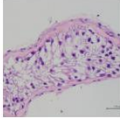
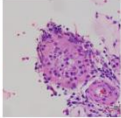
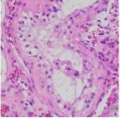
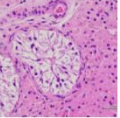
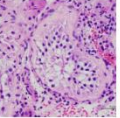
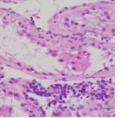


Google Cloud Platform Vision API interface showing evaluation results for the JSC_classifier model. The model is JSC_classifier_20200715041615 with a confidence threshold of 0.5.

Filter labels:

Label	Score
All labels	0.82611
JSC_1_2_3	0.94445
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False negatives: Your model should have predicted JSC_1_2_3 on these images.

Image	Score(s)
	0.4955174
	0.4995333
	0.49501184
	0.4766123
	0.48730017
	0.487054

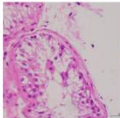
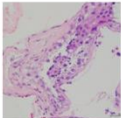
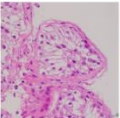
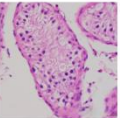
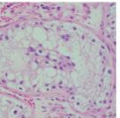
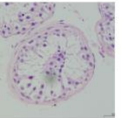
False positives: Your model incorrectly predicted JSC_1_2_3 on these images.

Google Cloud Platform Vision API interface showing evaluation results for the JSC_classifier model. The model is JSC_classifier_20200715041615 with a confidence threshold of 0.5.

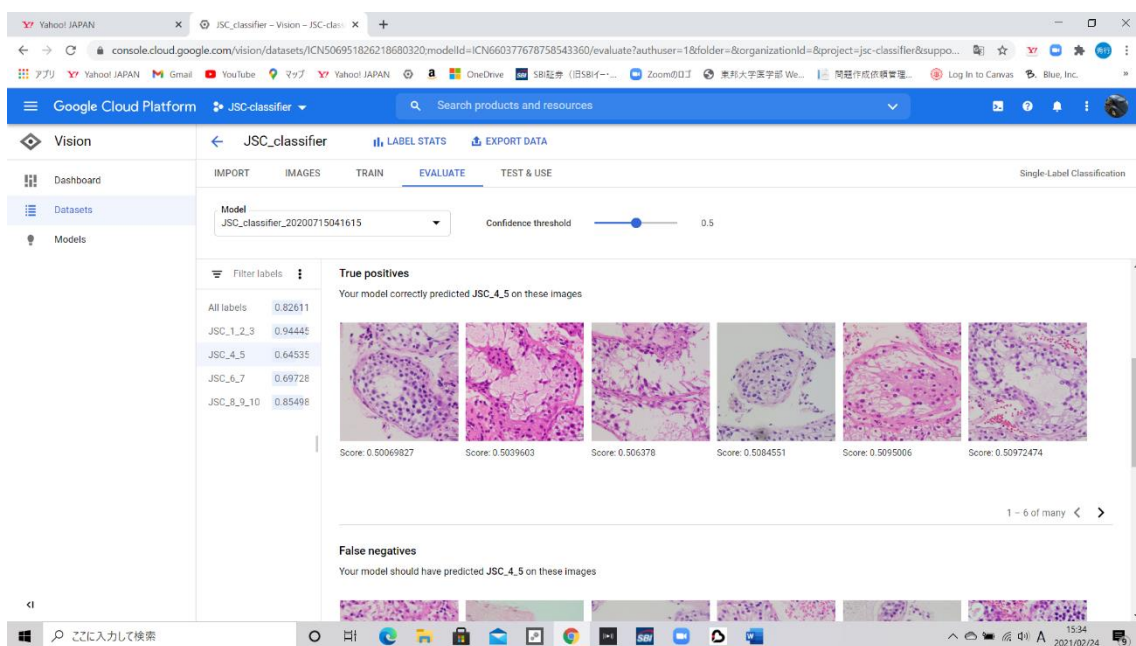
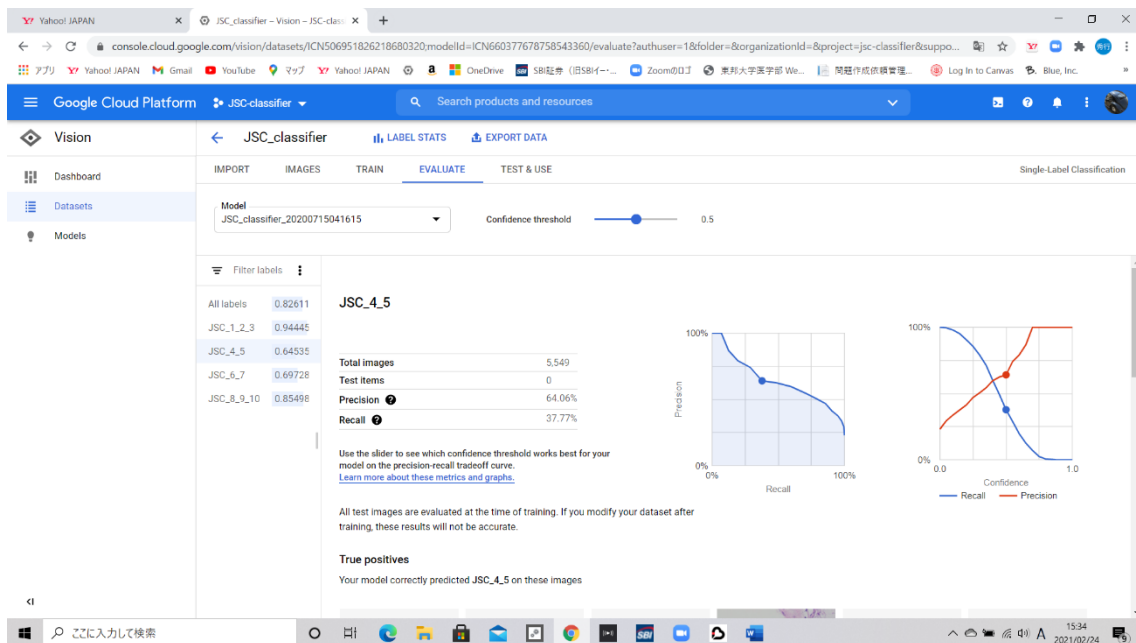
Filter labels:

Label	Score
All labels	0.82611
JSC_1_2_3	0.94445
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False positives: Your model incorrectly predicted JSC_1_2_3 on these images.

Image	Score
	0.506111
	0.50629357
	0.5229429
	0.5436324
	0.5495402
	0.5508962

The precision was 64.06% and recall was 37.77% for a Johnsen score of 4-5. True positive, false negative, and false positive images are indicated.



Google Cloud Platform Vision API interface for JSC_classifier. The model selected is JSC_classifier_20200715041615. The confidence threshold is set to 0.5. The interface shows a list of labels on the left and a 'False negatives' section on the right. The 'False negatives' section displays six histology images that the model incorrectly classified as JSC_4.5. Below the images, the scores for each are listed: 0.49015993, 0.49117517, 0.49306914, 0.4974781, 0.49106362, and 0.49746317.

Label	Score
All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False negatives
Your model should have predicted JSC_4_5 on these images

Score(s): 0.49015993 Score(s): 0.49117517 Score(s): 0.49306914 Score(s): 0.4974781 Score(s): 0.49106362 Score(s): 0.49746317

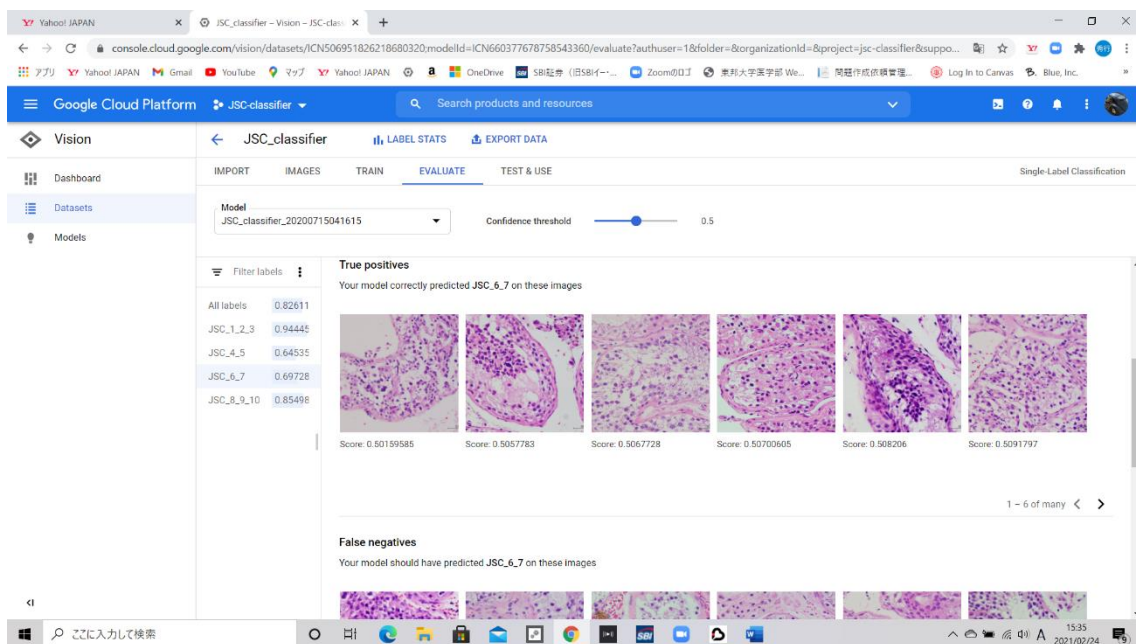
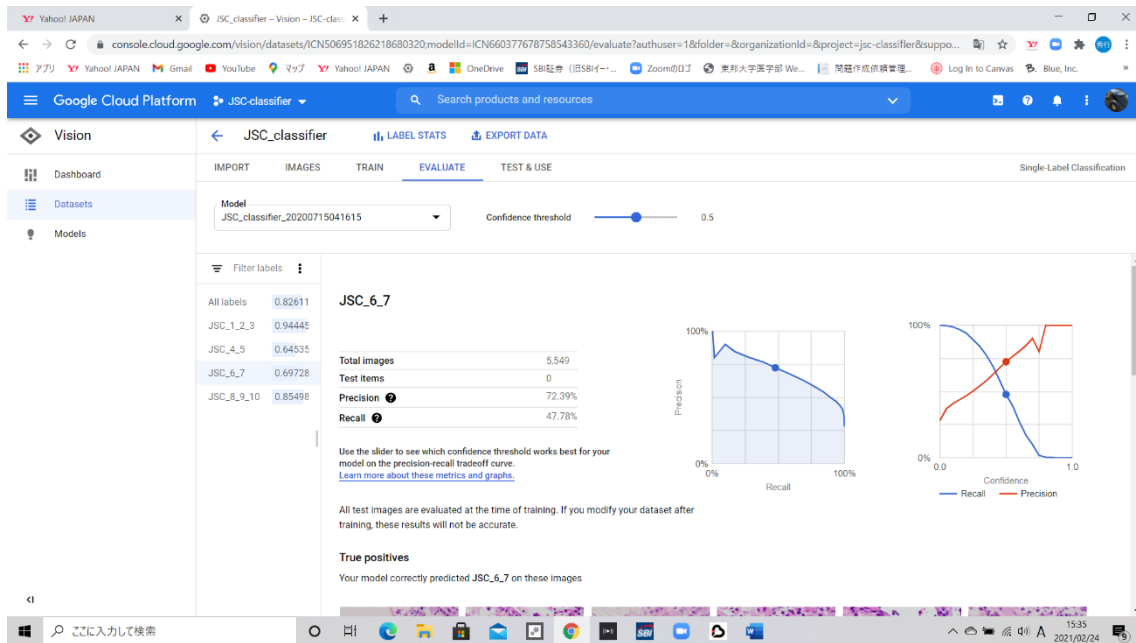
Google Cloud Platform Vision API interface for JSC_classifier. The model selected is JSC_classifier_20200715041615. The confidence threshold is set to 0.5. The interface shows a list of labels on the left and a 'False positives' section on the right. The 'False positives' section displays six histology images that the model incorrectly classified as JSC_4.5. Below the images, the scores for each are listed: 0.5005101, 0.50070107, 0.50211173, 0.50391006, 0.50409645, and 0.5042993.

Label	Score
All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False positives
Your model incorrectly predicted JSC_4_5 on these images

Score: 0.5005101 Score: 0.50070107 Score: 0.50211173 Score: 0.50391006 Score: 0.50409645 Score: 0.5042993

The precision was 72.39% and recall was 47.78% for a Johnsen score of 6-7. True positive, false negative, and false positive images are indicated.



Google Cloud Platform Vision API interface for JSC_classifier. The interface shows the 'EVALUATE' tab with a confidence threshold of 0.5. The 'Filter labels' section lists the following labels and their scores:

Label	Score
All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

The 'False negatives' section displays six histology images that the model incorrectly classified as JSC_6_7. The scores for these images are:

- Score(s): 0.49470612
- Score(s): 0.49559996
- Score(s): 0.4951952
- Score(s): 0.49427426
- Score(s): 0.49359143
- Score(s): 0.49630618

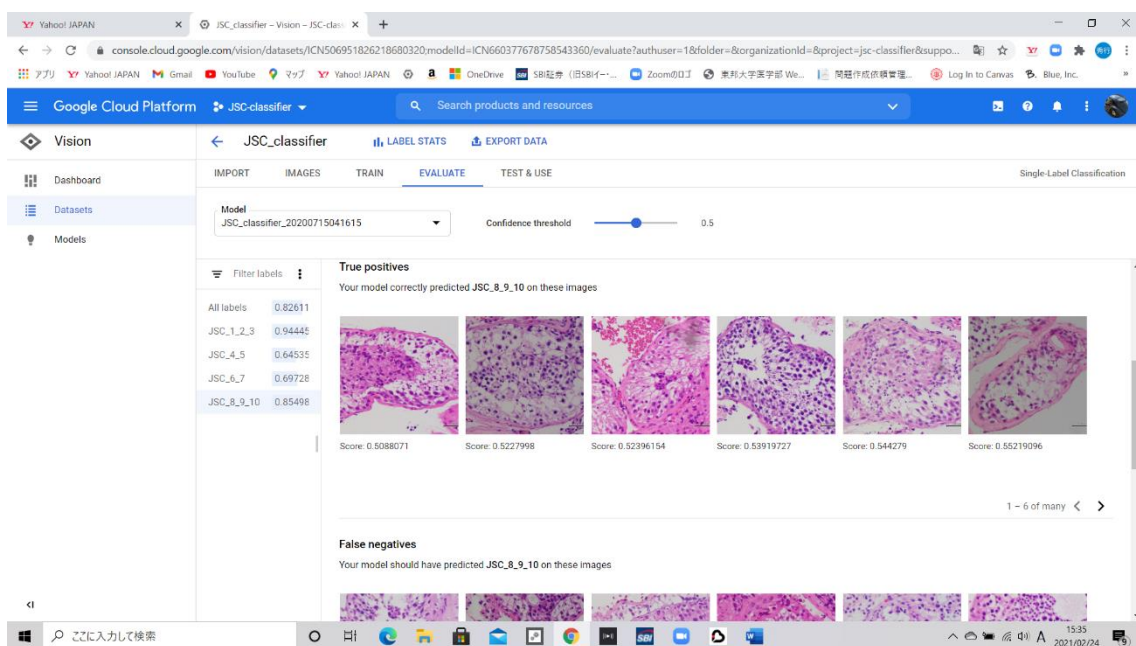
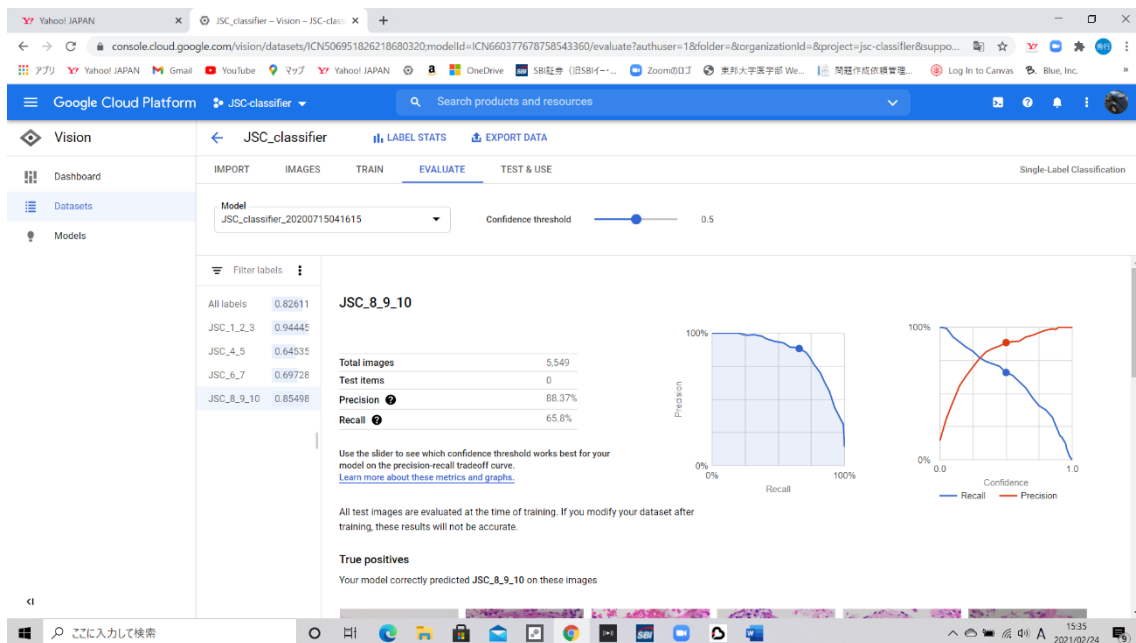
Google Cloud Platform Vision API interface for JSC_classifier. The interface shows the 'EVALUATE' tab with a confidence threshold of 0.5. The 'Filter labels' section lists the following labels and their scores:

Label	Score
All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

The 'False positives' section displays six histology images that the model incorrectly classified as JSC_6_7. The scores for these images are:

- Score: 0.5000983
- Score: 0.50214577
- Score: 0.5065933
- Score: 0.507078
- Score: 0.507088
- Score: 0.51084286

The precision was 88.37% and recall was 65.8% for a Johnsen score of 8-10. True positive, false negative, and false positive images are indicated.

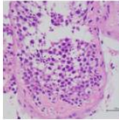
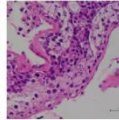
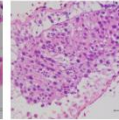
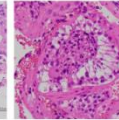
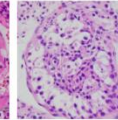
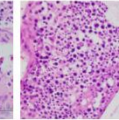


Google Cloud Platform Vision API interface for JSC_classifier. The model is JSC_classifier_20200715041615. The confidence threshold is set to 0.5. The interface shows a list of labels on the left and a grid of images under the 'False negatives' section.

Filter labels:

All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False negatives: Your model should have predicted JSC_8_9_10 on these images.

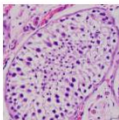
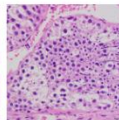
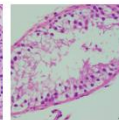
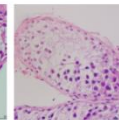
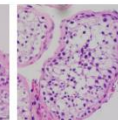
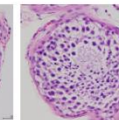
					
Score(s): 0.49940252	Score(s): 0.49946204	Score(s): 0.48098984	Score(s): 0.49711955	Score(s): 0.47661543	Score(s): 0.49048218

Google Cloud Platform Vision API interface for JSC_classifier. The model is JSC_classifier_20200715041615. The confidence threshold is set to 0.5. The interface shows a list of labels on the left and a grid of images under the 'False positives' section.

Filter labels:

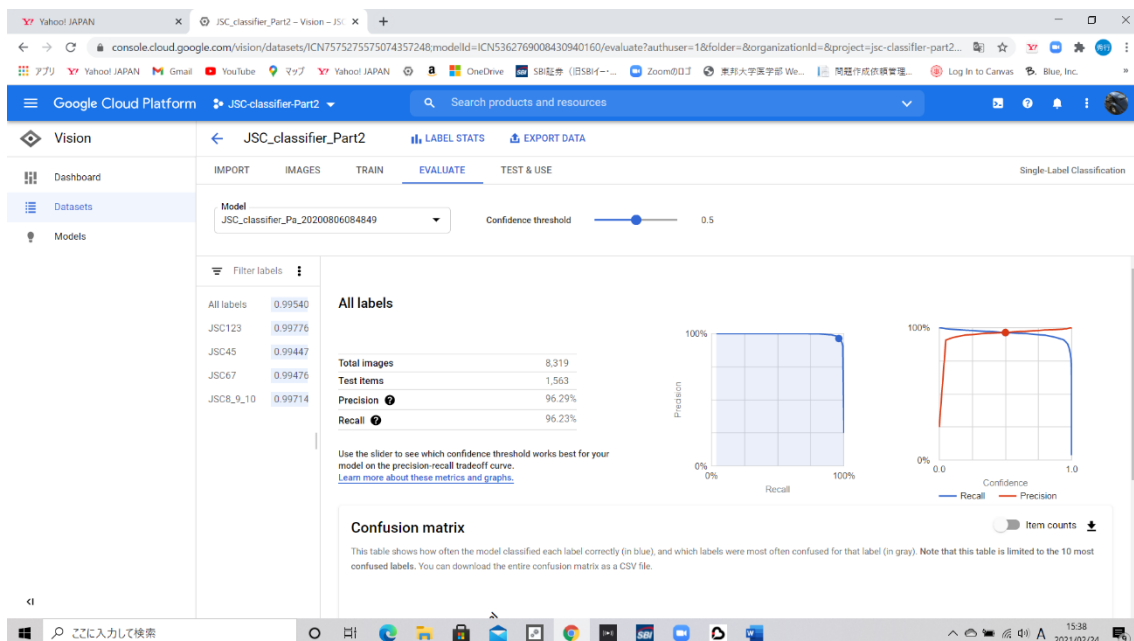
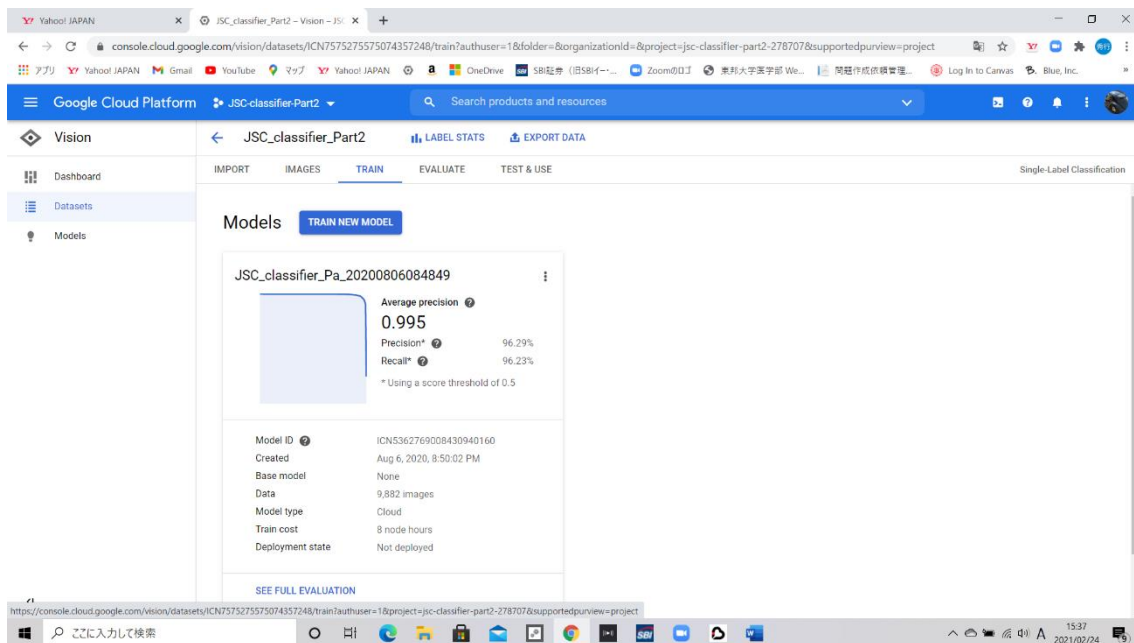
All labels	0.82611
JSC_1_2_3	0.94444
JSC_4_5	0.64535
JSC_6_7	0.69728
JSC_8_9_10	0.85498

False positives: Your model incorrectly predicted JSC_8_9_10 on these images.

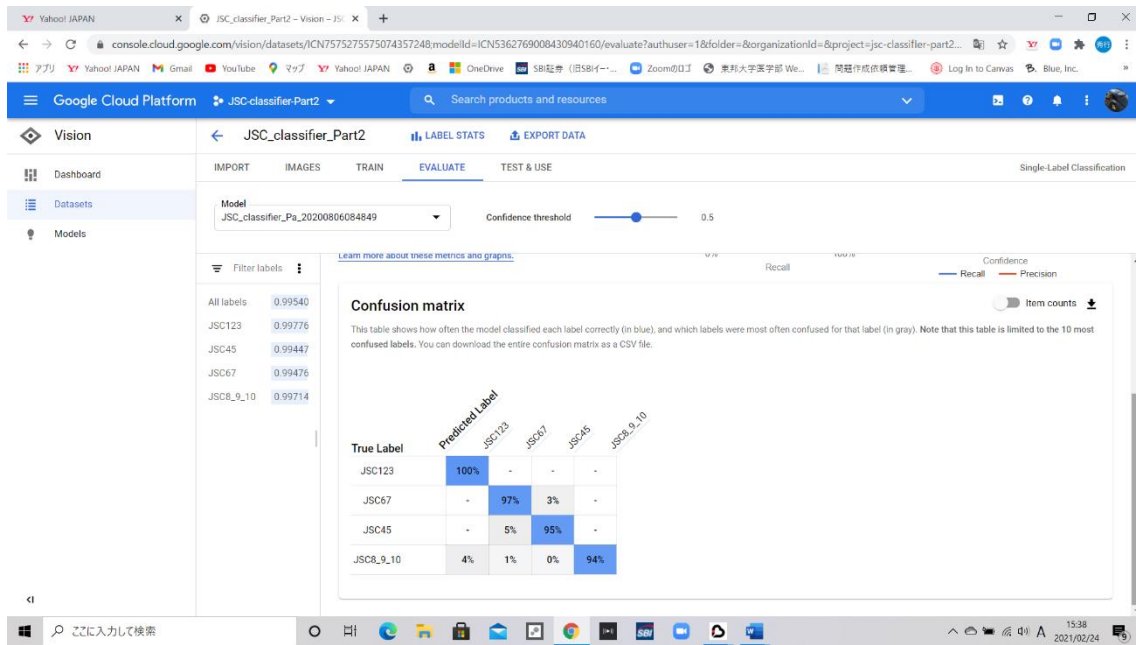
					
Score: 0.50613654	Score: 0.5077303	Score: 0.5599204	Score: 0.58292073	Score: 0.6054845	Score: 0.61686766

JSC_classifier_Part 2: image dataset of 9822 expansion images (5.0 X 5.0 cm)

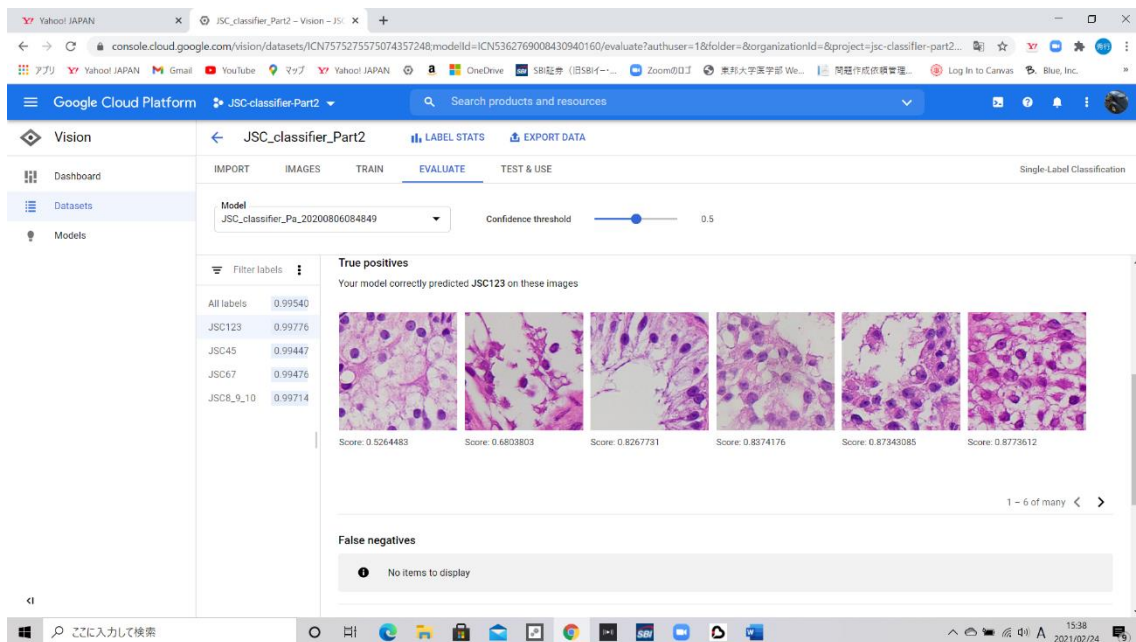
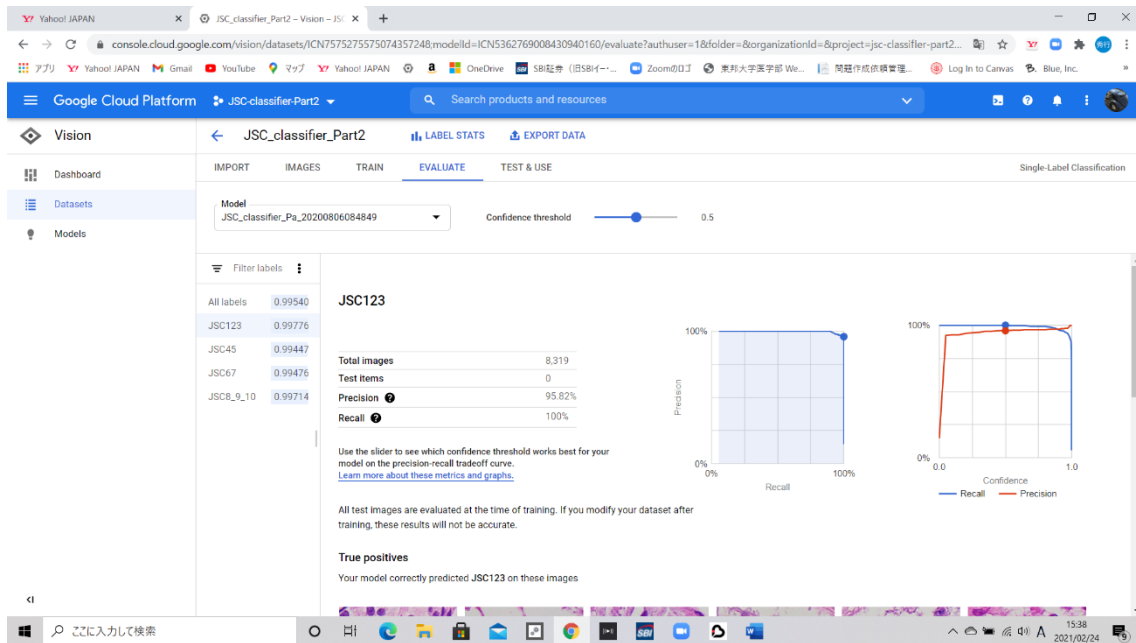
The average precision (positive predictive value) of the algorithm was 99.5%, precision was 96.29%, and recall was 96.23% for all labels.



Confusion matrix



The precision was 95.82% and recall was 100% for a Johnsen score of 1-3. True positive and false positive images are indicated.



Google Cloud Platform Vision API Evaluation Page for JSC_classifier_Part2.

Model: JSC_classifier_Pa_20200806084849

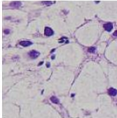
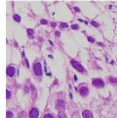
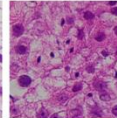
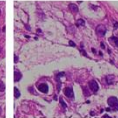
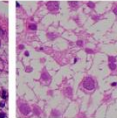
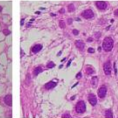
Confidence threshold: 0.5

Filter Labels:

Label	Score
All labels	0.99540
JSC123	0.99776
JSC45	0.99447
JSC67	0.99476
JSC8_9_10	0.99714

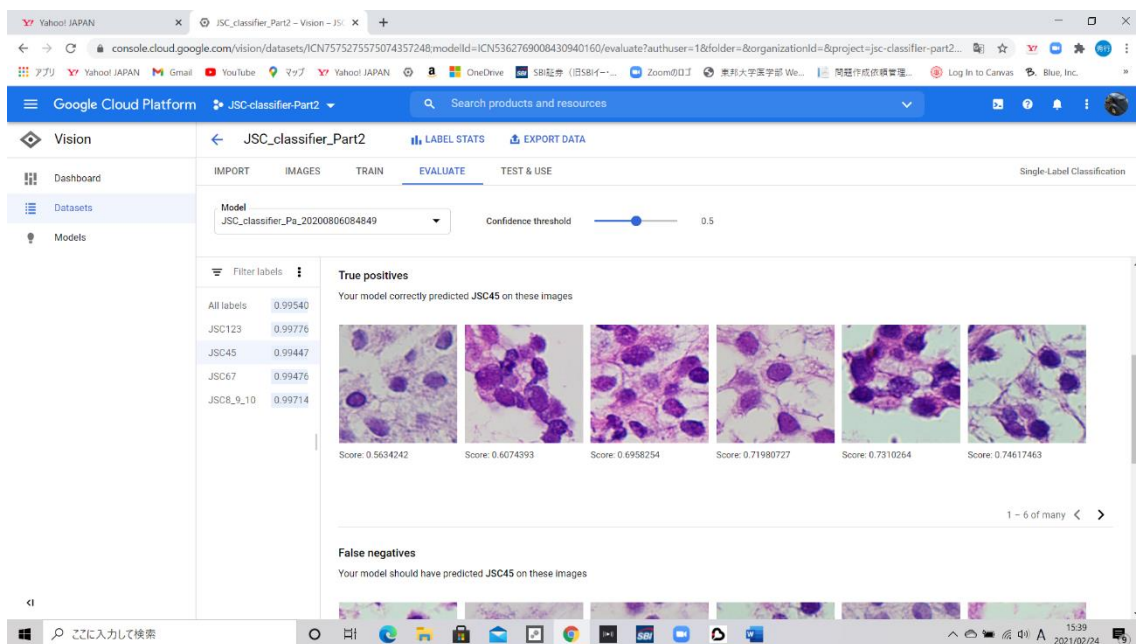
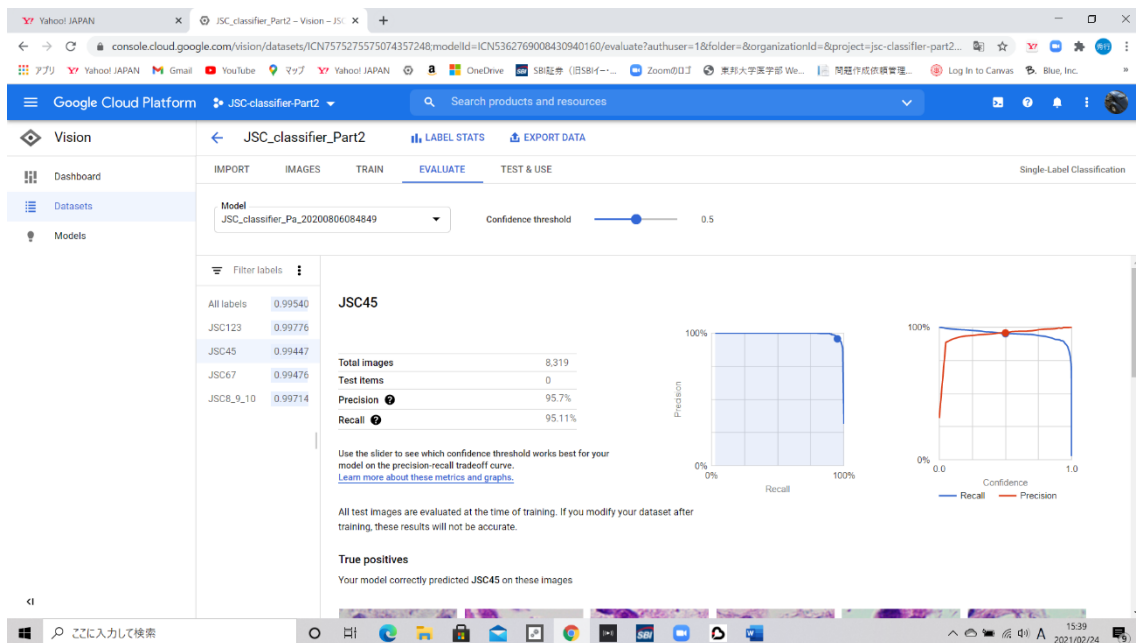
False negatives: No items to display

False positives: Your model incorrectly predicted JSC123 on these images

Image	Score
	0.51657626
	0.60036254
	0.8062527
	0.91387177
	0.94386536
	0.97502346

1 - 6 of many

The precision was 95.7% and recall was 95.11% for a Johnsen score of 4-5. True positive, false negative, and false positive images are indicated.



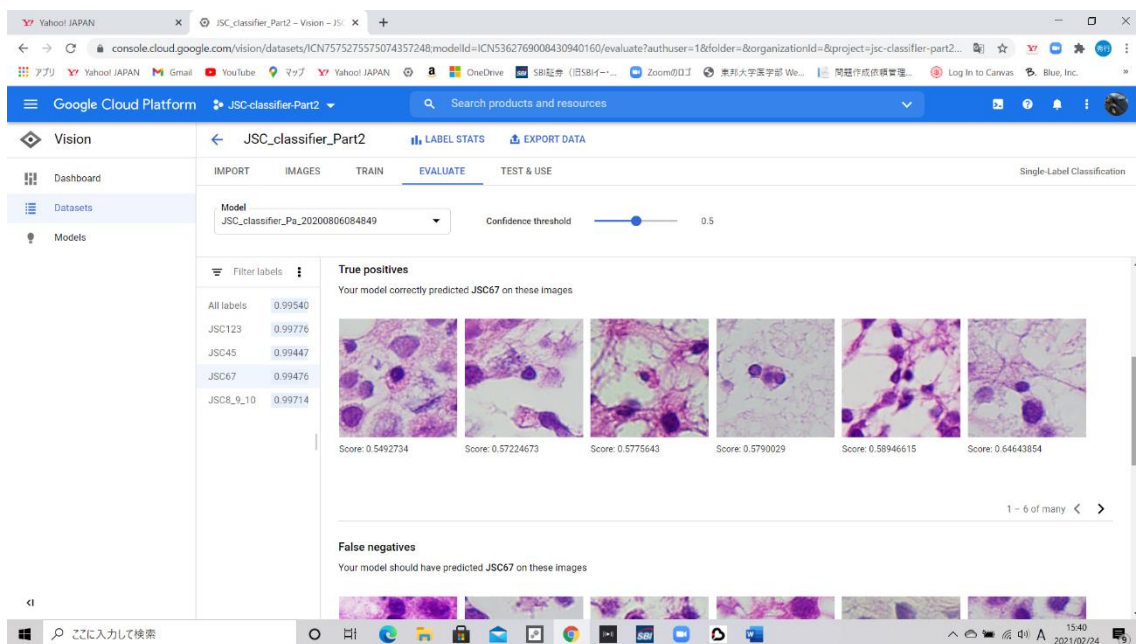
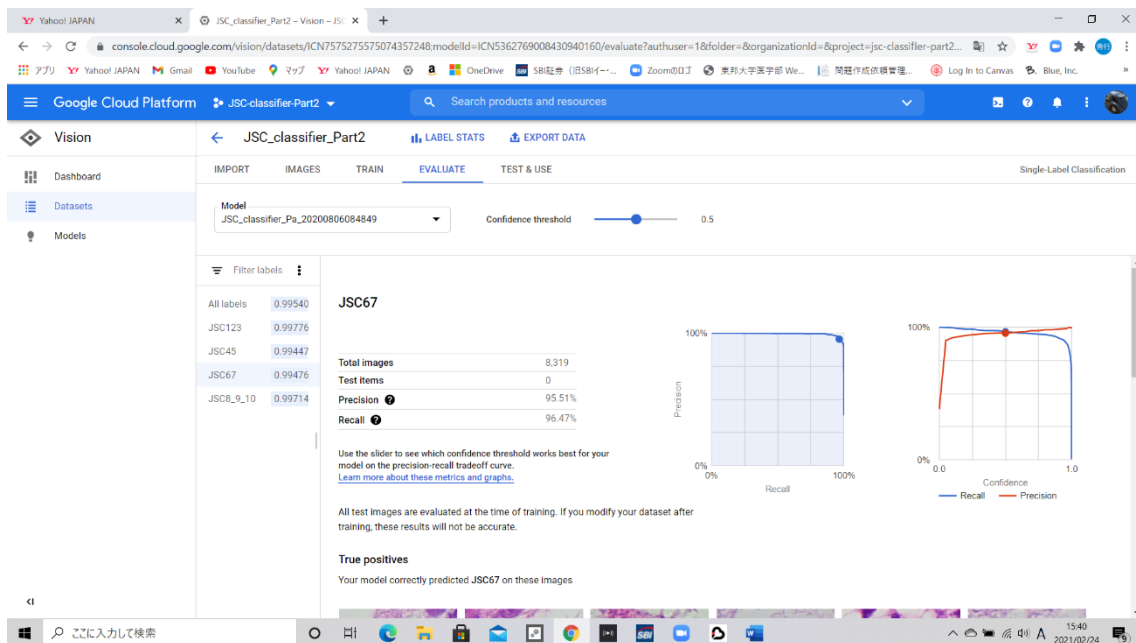
Google Cloud Platform Vision API Evaluation Page (Screenshot 1):

- Model:** JSC_classifier_Pa_20200806084849
- Confidence threshold:** 0.5
- Filter labels:**
 - All labels: 0.99540
 - JSC123: 0.99776
 - JSC45: 0.99447
 - JSC67: 0.99476
 - JSC8_9_10: 0.99714
- False negatives:** Your model should have predicted JSC45 on these images.
 - Image 1: Score(s): 0.41148195
 - Image 2: Score(s): 0.45436364
 - Image 3: Score(s): 0.47549698
 - Image 4: Score(s): 0.41683275
 - Image 5: Score(s): 0.44331735
 - Image 6: Score(s): 0.40800672
- False positives:** Your model incorrectly predicted JSC45 on these images. (Images are visible but scores are not clearly legible).

Google Cloud Platform Vision API Evaluation Page (Screenshot 2):

- Model:** JSC_classifier_Pa_20200806084849
- Confidence threshold:** 0.5
- Filter labels:**
 - All labels: 0.99540
 - JSC123: 0.99776
 - JSC45: 0.99447
 - JSC67: 0.99476
 - JSC8_9_10: 0.99714
- False positives:** Your model incorrectly predicted JSC45 on these images.
 - Image 1: Score: 0.5014175
 - Image 2: Score: 0.5203256
 - Image 3: Score: 0.52046686
 - Image 4: Score: 0.5414395
 - Image 5: Score: 0.5442471
 - Image 6: Score: 0.5719388

The precision was 95.51% and recall was 96.47% for a Johnsen score of 6-7. True positive, false negative, and false positive images are indicated.



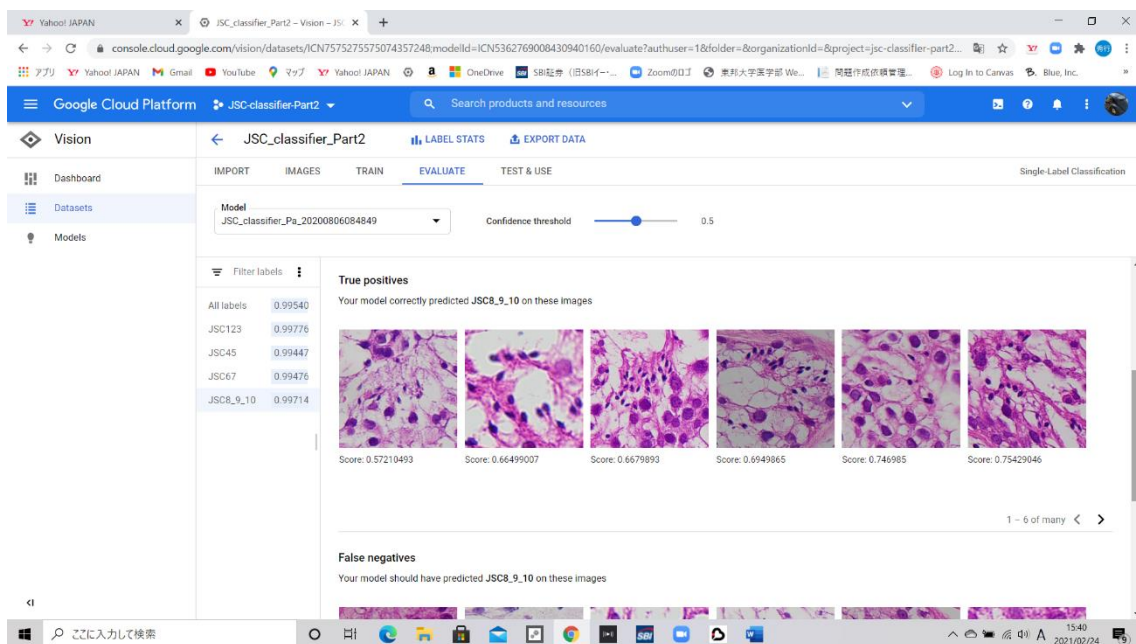
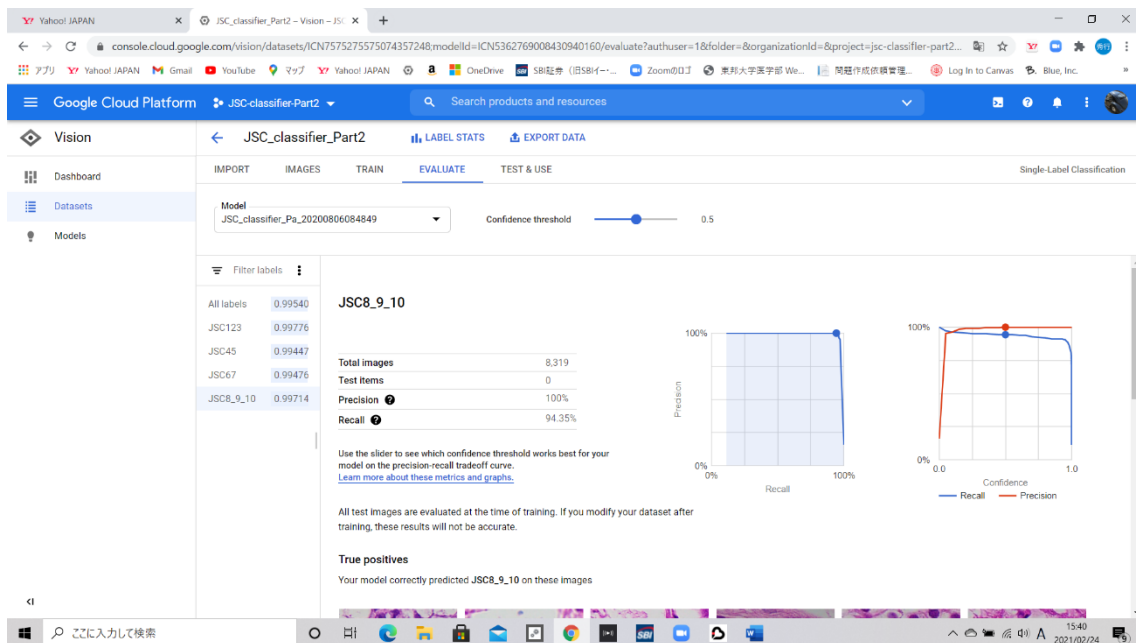
Google Cloud Platform Vision API Evaluation Page (Screenshot 1):

- Model:** JSC_classifier_Pa_20200806084849
- Confidence threshold:** 0.5
- Filter labels:**
 - All labels: 0.99540
 - JSC123: 0.99776
 - JSC45: 0.99447
 - JSC67: 0.99476
 - JSC8_9_10: 0.99714
- False negatives:** Your model should have predicted JSC67 on these images.
 - Image 1: Score(s): 0.49531528
 - Image 2: Score(s): 0.45855585
 - Image 3: Score(s): 0.47967267
 - Image 4: Score(s): 0.42796662
 - Image 5: Score(s): 0.47953302
 - Image 6: Score(s): 0.49857822
- False positives:** Your model incorrectly predicted JSC67 on these images. (Images are visible but not clearly legible in this view)

Google Cloud Platform Vision API Evaluation Page (Screenshot 2):

- Model:** JSC_classifier_Pa_20200806084849
- Confidence threshold:** 0.5
- Filter labels:**
 - All labels: 0.99540
 - JSC123: 0.99776
 - JSC45: 0.99447
 - JSC67: 0.99476
 - JSC8_9_10: 0.99714
- False positives:** Your model incorrectly predicted JSC67 on these images.
 - Image 1: Score: 0.524503
 - Image 2: Score: 0.54563326
 - Image 3: Score: 0.55668014
 - Image 4: Score: 0.583166
 - Image 5: Score: 0.58824676
 - Image 6: Score: 0.5919871

The precision was 100% and recall was 94.35% for a Johnsen score of 8-10. True positive and false negative images are indicated.



Google Cloud Platform Vision API interface for JSC_classifier_Part2. The interface shows the evaluation results for a model, specifically focusing on false negatives.

Model: JSC_classifier_Pa_20200806084849
Confidence threshold: 0.5

Filter Labels:

Label	Score(s)
All labels	0.99540
JSC123	0.99776
JSC45	0.99447
JSC67	0.99476
JSC8_9_10	0.99714

False negatives:
 Your model should have predicted JSC8_9_10 on these images

6 images are displayed with their respective scores:

Image	Score(s)
	0.19374532
	0.42604476
	0.21499024
	0.39963624
	0.08526694
	0.13940889

1 - 6 of many

False positives:
 No items to display