

2018 Model Ensembles Evaluation Using 2020 Training Data and 2018 Validation Data

2020 Labels	Count
unknown	27124
nevus	5193
melanoma	584
seborrheic keratosis	135
lentigo NOS	44
lichenoid keratosis	37
solar lentigo	7
cafe-au-lait macule	1
atypical melanocytic proliferation	1
Total	33126

After removing duplicates

2020 Labels	Count
unknown	26706
nevus	5191
melanoma	581
seborrheic keratosis	135
lentigo NOS	44
lichenoid keratosis	35
solar lentigo	7
cafe-au-lait macule	1
atypical melanocytic proliferation	1
Total	32701

The Tables above shows the number of images that are in each of the 2020 classes both before and after removing the duplicate images.

Mapping of labels between 2018 and 2020 Data Set

As we can see from the diagnosis labels of 2020, there does not exist a direct match of all labels between that of 2020 and 2018. Only ‘melanoma’ and ‘nevus’ can be directly matched. Consequently, some of the 2020 labels is mapped according to the data description of the 2019 data set.

Label Index	2018 Labels	2020 Labels
0	AKIEC	
1	BCC	
2	BKL	seborrheic keratosis ¹ solar lentigo ¹ lichenoid keratosis ¹ lentigo NOS ¹
3	DF	
4	MEL	melanoma
5	NV	nevus
6	VASC	
		cafe-au-lait macule Atypical melanocytic proliferation Unknown

¹ mapped according to description in <https://challenge2019.isic-archive.com>

Summary of Performance of the 3 ensembles

ensemble	Ensemble Description	2018 Validation Data Set Overall Accuracy	2020 'Training' Data Set Overall Accuracy
1	Ensemble of Base Model	74%	84 %
2	Ensemble of Modified Base Model (Dropout 25%, 1 extra CNN Layer)	2%	3%
3	Ensemble of Modified Base Model (Dropout 40%, 1 extra CNN Layer)	77%	83%

In summary, Ensemble 2 did not provide an optimal performance on the prediction of both the 2018 validation set and the 2020 data set. In terms of the 2020 data set and the 2018 validation data set, ensemble 1 and ensemble 3 similar performance. Ensemble 3 performed relatively better in the 2018 validation as there more classes in the data thus, is less skewed compared to the 2020 data set. In both the data set used in the prediction, we can see a high recall and precision for the nevus class. The tables below shows the detailed result of the predictions and samples of images that were correctly and incorrectly classified.

2020 Data Prediction (bkl, nv, mel)

Ensemble	Result
1	<pre> [[0 0 0 0 0 0 0] [0 0 0 0 0 0 0] [13 25 11 7 7 157 1] [0 0 0 0 0 0 0] [10 32 17 2 69 449 2] [14 58 49 22 59 4976 13] [0 0 0 0 0 0 0]] precision recall f1-score support akiel 0.00 0.00 0.00 0 bcc 0.00 0.00 0.00 0 bkl 0.14 0.05 0.07 221 df 0.00 0.00 0.00 0 mel 0.51 0.12 0.19 581 nv 0.89 0.96 0.92 5191 vasc 0.00 0.00 0.00 0 accuracy 0.84 5993 macro avg 0.22 0.16 0.17 5993 weighted avg 0.83 0.84 0.82 5993 </pre>
2	<pre> [[141 80 0 0] [0 0 0 0] [205 376 0 0] [1250 3941 0 0]] precision recall f1-score support bkl 0.09 0.64 0.16 221 df 0.00 0.00 0.00 0 mel 0.00 0.00 0.00 581 nv 0.00 0.00 0.00 5191 accuracy 0.02 5993 macro avg 0.02 0.16 0.04 5993 weighted avg 0.00 0.02 0.01 5993 </pre>
3	<pre> ensemble 3 [[0 0 0 0 0 0 0] [0 0 0 0 0 0 0] [27 27 28 16 13 110 0] [0 0 0 0 0 0 0] [31 34 43 10 78 382 3] [50 67 79 31 93 4856 15] [0 0 0 0 0 0 0]] precision recall f1-score support 0 0.00 0.00 0.00 0 1 0.00 0.00 0.00 0 2 0.19 0.13 0.15 221 3 0.00 0.00 0.00 0 4 0.42 0.13 0.20 581 5 0.91 0.94 0.92 5191 6 0.00 0.00 0.00 0 accuracy 0.83 5993 macro avg 0.22 0.17 0.18 5993 weighted avg 0.83 0.83 0.82 5993 </pre>

2020 Data Prediction (nv and mel)

Ensemble																																																			
1	<p>ensemble 1 only nv and mel</p> <pre>[[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 0 0] [10 32 17 2 69 449 2] [14 58 49 22 59 4976 13] [0 0 0 0 0 0 0]] precision recall f1-score support</pre> <table> <tbody> <tr><td>0</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>1</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>2</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>3</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>4</td><td>0.54</td><td>0.12</td><td>0.19</td><td>581</td></tr> <tr><td>5</td><td>0.92</td><td>0.96</td><td>0.94</td><td>5191</td></tr> <tr><td>6</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.87</td><td>5772</td></tr> <tr><td>macro avg</td><td>0.21</td><td>0.15</td><td>0.16</td><td>5772</td></tr> <tr><td>weighted avg</td><td>0.88</td><td>0.87</td><td>0.86</td><td>5772</td></tr> </tbody> </table>	0	0.00	0.00	0.00	0	1	0.00	0.00	0.00	0	2	0.00	0.00	0.00	0	3	0.00	0.00	0.00	0	4	0.54	0.12	0.19	581	5	0.92	0.96	0.94	5191	6	0.00	0.00	0.00	0	accuracy			0.87	5772	macro avg	0.21	0.15	0.16	5772	weighted avg	0.88	0.87	0.86	5772
0	0.00	0.00	0.00	0																																															
1	0.00	0.00	0.00	0																																															
2	0.00	0.00	0.00	0																																															
3	0.00	0.00	0.00	0																																															
4	0.54	0.12	0.19	581																																															
5	0.92	0.96	0.94	5191																																															
6	0.00	0.00	0.00	0																																															
accuracy			0.87	5772																																															
macro avg	0.21	0.15	0.16	5772																																															
weighted avg	0.88	0.87	0.86	5772																																															
2	<p>ensemble 2 only nv and mel</p> <pre>[[0 0 0 0] [0 0 0 0] [205 376 0 0] [1250 3941 0 0]] precision recall f1-score support</pre> <table> <tbody> <tr><td>2</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.0</td></tr> <tr><td>3</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.0</td></tr> <tr><td>4</td><td>0.00</td><td>0.00</td><td>0.00</td><td>581.0</td></tr> <tr><td>5</td><td>0.00</td><td>0.00</td><td>0.00</td><td>5191.0</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.00</td><td>5772.0</td></tr> <tr><td>macro avg</td><td>0.00</td><td>0.00</td><td>0.00</td><td>5772.0</td></tr> <tr><td>weighted avg</td><td>0.00</td><td>0.00</td><td>0.00</td><td>5772.0</td></tr> </tbody> </table>	2	0.00	0.00	0.00	0.0	3	0.00	0.00	0.00	0.0	4	0.00	0.00	0.00	581.0	5	0.00	0.00	0.00	5191.0	accuracy			0.00	5772.0	macro avg	0.00	0.00	0.00	5772.0	weighted avg	0.00	0.00	0.00	5772.0															
2	0.00	0.00	0.00	0.0																																															
3	0.00	0.00	0.00	0.0																																															
4	0.00	0.00	0.00	581.0																																															
5	0.00	0.00	0.00	5191.0																																															
accuracy			0.00	5772.0																																															
macro avg	0.00	0.00	0.00	5772.0																																															
weighted avg	0.00	0.00	0.00	5772.0																																															
3	<p>ensemble 3</p> <pre>[[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 0 0] [31 34 43 10 78 382 3] [50 67 79 31 93 4856 15] [0 0 0 0 0 0 0]] precision recall f1-score support</pre> <table> <tbody> <tr><td>0</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>1</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>2</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>3</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>4</td><td>0.46</td><td>0.13</td><td>0.21</td><td>581</td></tr> <tr><td>5</td><td>0.93</td><td>0.94</td><td>0.93</td><td>5191</td></tr> <tr><td>6</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.85</td><td>5772</td></tr> <tr><td>macro avg</td><td>0.20</td><td>0.15</td><td>0.16</td><td>5772</td></tr> <tr><td>weighted avg</td><td>0.88</td><td>0.85</td><td>0.86</td><td>5772</td></tr> </tbody> </table>	0	0.00	0.00	0.00	0	1	0.00	0.00	0.00	0	2	0.00	0.00	0.00	0	3	0.00	0.00	0.00	0	4	0.46	0.13	0.21	581	5	0.93	0.94	0.93	5191	6	0.00	0.00	0.00	0	accuracy			0.85	5772	macro avg	0.20	0.15	0.16	5772	weighted avg	0.88	0.85	0.86	5772
0	0.00	0.00	0.00	0																																															
1	0.00	0.00	0.00	0																																															
2	0.00	0.00	0.00	0																																															
3	0.00	0.00	0.00	0																																															
4	0.46	0.13	0.21	581																																															
5	0.93	0.94	0.93	5191																																															
6	0.00	0.00	0.00	0																																															
accuracy			0.85	5772																																															
macro avg	0.20	0.15	0.16	5772																																															
weighted avg	0.88	0.85	0.86	5772																																															

2018 Validation Data Prediction

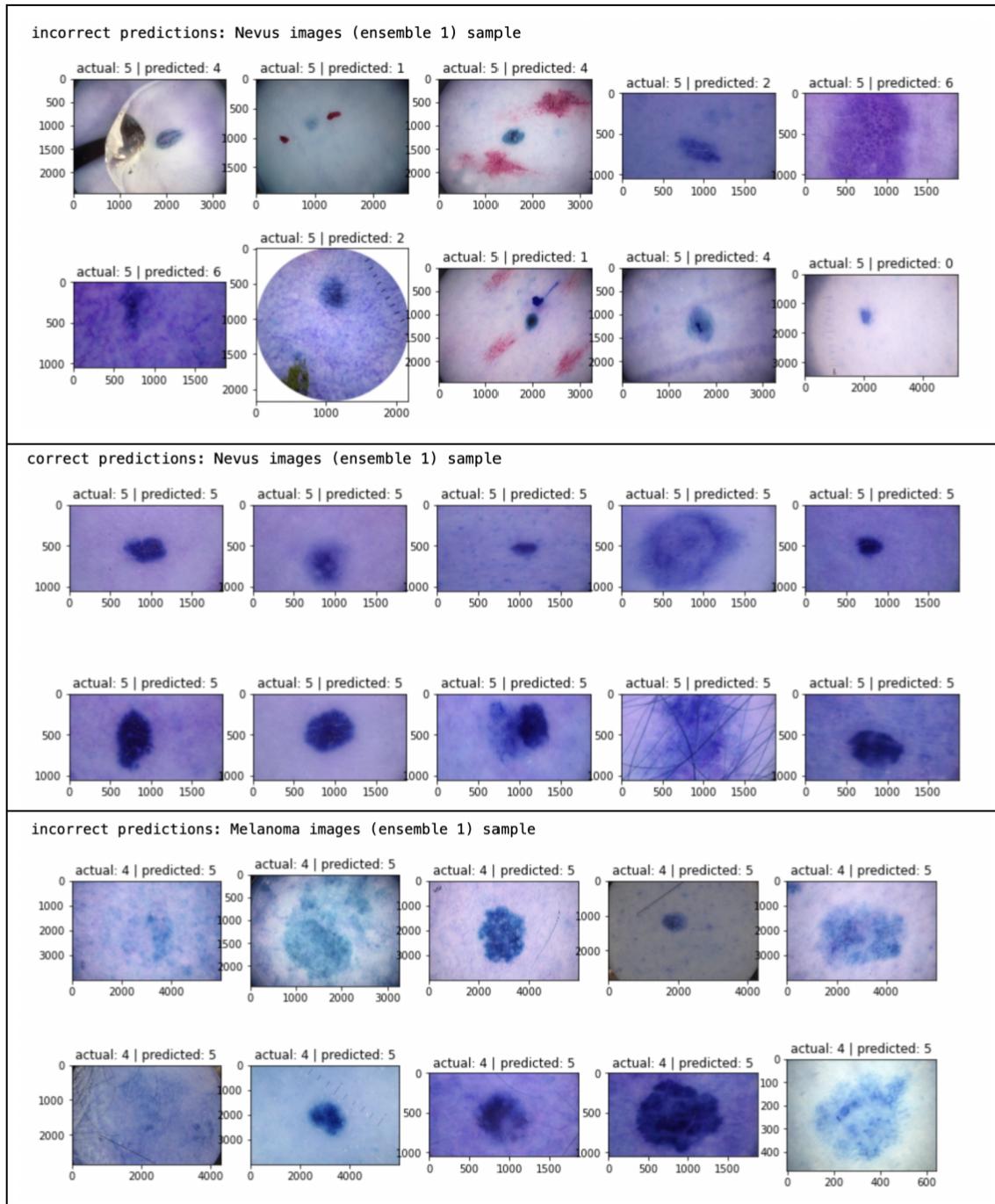
Ensemble																																																									
1	<pre> Confusion Matrix [[3 1 1 0 0 3 0] [1 11 0 0 0 3 0] [0 0 7 0 1 14 0] [0 0 1 0 0 0 0] [0 0 2 0 5 14 0] [0 0 4 1 2 116 0] [0 1 0 0 0 1 1]]</pre> <table> <thead> <tr> <th colspan="2">Classification Report</th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>akiec</td> <td>0.75</td> <td>0.38</td> <td>0.50</td> <td>8</td> </tr> <tr> <td>bcc</td> <td>0.85</td> <td>0.73</td> <td>0.79</td> <td>15</td> </tr> <tr> <td>bkl</td> <td>0.47</td> <td>0.32</td> <td>0.38</td> <td>22</td> </tr> <tr> <td>df</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>1</td> </tr> <tr> <td>mel</td> <td>0.62</td> <td>0.24</td> <td>0.34</td> <td>21</td> </tr> <tr> <td>nv</td> <td>0.77</td> <td>0.94</td> <td>0.85</td> <td>123</td> </tr> <tr> <td>vasc</td> <td>1.00</td> <td>0.33</td> <td>0.50</td> <td>3</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.74</td> <td>193</td> </tr> <tr> <td>macro avg</td> <td>0.64</td> <td>0.42</td> <td>0.48</td> <td>193</td> </tr> <tr> <td>weighted avg</td> <td>0.72</td> <td>0.74</td> <td>0.71</td> <td>193</td> </tr> </tbody> </table>	Classification Report		precision	recall	f1-score	support	akiec	0.75	0.38	0.50	8	bcc	0.85	0.73	0.79	15	bkl	0.47	0.32	0.38	22	df	0.00	0.00	0.00	1	mel	0.62	0.24	0.34	21	nv	0.77	0.94	0.85	123	vasc	1.00	0.33	0.50	3	accuracy			0.74	193	macro avg	0.64	0.42	0.48	193	weighted avg	0.72	0.74	0.71	193
Classification Report		precision	recall	f1-score	support																																																				
akiec	0.75	0.38	0.50	8																																																					
bcc	0.85	0.73	0.79	15																																																					
bkl	0.47	0.32	0.38	22																																																					
df	0.00	0.00	0.00	1																																																					
mel	0.62	0.24	0.34	21																																																					
nv	0.77	0.94	0.85	123																																																					
vasc	1.00	0.33	0.50	3																																																					
accuracy			0.74	193																																																					
macro avg	0.64	0.42	0.48	193																																																					
weighted avg	0.72	0.74	0.71	193																																																					
2	<pre> Confusion Matrix [[0 0 1 7 0 0 0] [0 0 7 8 0 0 0] [0 0 6 16 0 0 0] [0 0 1 0 0 0 0] [0 0 2 19 0 0 0] [0 0 22 101 0 0 0] [0 0 0 3 0 0 0]]</pre> <table> <thead> <tr> <th colspan="2">Classification Report</th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>akiec</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>8</td> </tr> <tr> <td>bcc</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>15</td> </tr> <tr> <td>bkl</td> <td>0.15</td> <td>0.27</td> <td>0.20</td> <td>22</td> </tr> <tr> <td>df</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>1</td> </tr> <tr> <td>mel</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>21</td> </tr> <tr> <td>nv</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>123</td> </tr> <tr> <td>vasc</td> <td>0.00</td> <td>0.00</td> <td>0.00</td> <td>3</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.03</td> <td>193</td> </tr> <tr> <td>macro avg</td> <td>0.02</td> <td>0.04</td> <td>0.03</td> <td>193</td> </tr> <tr> <td>weighted avg</td> <td>0.02</td> <td>0.03</td> <td>0.02</td> <td>193</td> </tr> </tbody> </table>	Classification Report		precision	recall	f1-score	support	akiec	0.00	0.00	0.00	8	bcc	0.00	0.00	0.00	15	bkl	0.15	0.27	0.20	22	df	0.00	0.00	0.00	1	mel	0.00	0.00	0.00	21	nv	0.00	0.00	0.00	123	vasc	0.00	0.00	0.00	3	accuracy			0.03	193	macro avg	0.02	0.04	0.03	193	weighted avg	0.02	0.03	0.02	193
Classification Report		precision	recall	f1-score	support																																																				
akiec	0.00	0.00	0.00	8																																																					
bcc	0.00	0.00	0.00	15																																																					
bkl	0.15	0.27	0.20	22																																																					
df	0.00	0.00	0.00	1																																																					
mel	0.00	0.00	0.00	21																																																					
nv	0.00	0.00	0.00	123																																																					
vasc	0.00	0.00	0.00	3																																																					
accuracy			0.03	193																																																					
macro avg	0.02	0.04	0.03	193																																																					
weighted avg	0.02	0.03	0.02	193																																																					

3

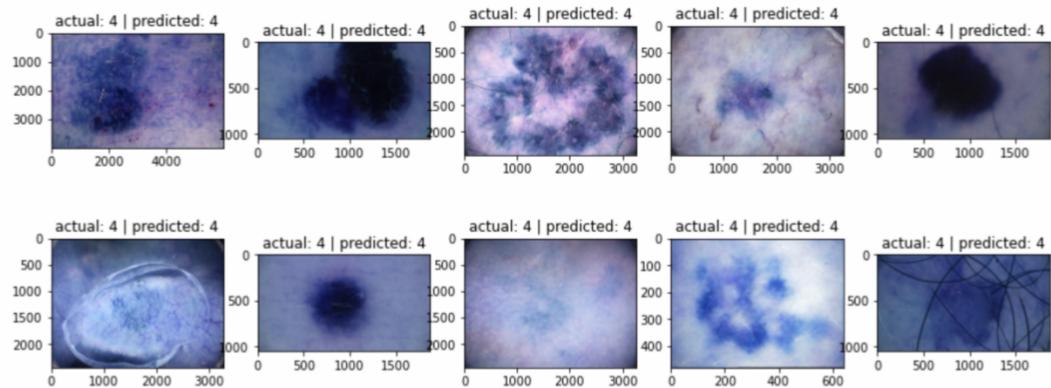
Confusion Matrix							
[6	0	1	0	0	1	0]
[2	13	0	0	0	0	0]
[2	1	13	0	3	3	0]
[0	0	1	0	0	0	0]
[0	1	4	0	12	4	0]
[0	2	10	1	7	103	0]
[0	2	0	0	0	0	1]]

Classification Report						
	precision	recall	f1-score	support		
akiec	0.60	0.75	0.67	8		
bcc	0.68	0.87	0.76	15		
bkl	0.45	0.59	0.51	22		
df	0.00	0.00	0.00	1		
mel	0.55	0.57	0.56	21		
nv	0.93	0.84	0.88	123		
vasc	1.00	0.33	0.50	3		
accuracy				0.77	193	
macro avg	0.60	0.56	0.55	193		
weighted avg	0.80	0.77	0.77	193		

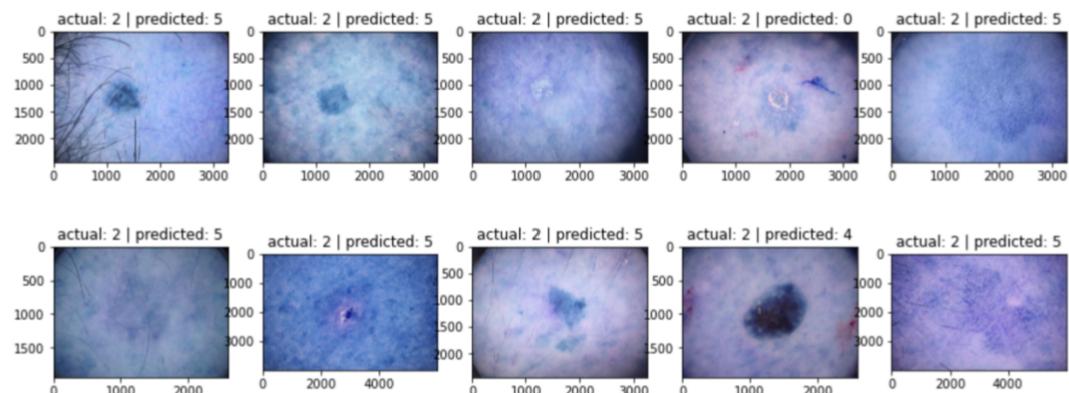
2020 Data Prediction result samples



correct predictions: Melanoma images (ensemble 1) sample



incorrect predictions: BKL images (ensemble 1) sample



correct predictions: BKL images (ensemble 1) sample

