

1 Flat Image Only Model

Table 1. BACC scores of the flat classifiers with only images and the number of parameters of each CNN. The cell highlighted in green represents the highest score.

CNN	BACC (%)	Number of parameters
ResNet50	70.04	23.59 M
ResNet101	68.56	42.66 M
DenseNet121	71.85	7.04 M
EfficientNet-B0	70.96	4.05 M
EfficientNet-B2	73.05	7.77 M

2 Hierarchical Models

2.1 Image Only Models

Table 2. BACC scores of the image only hierarchical classifiers. The cells highlighted in green represent the highest scores for each classifier excluding the model Combined 0.

CNN\Classifier	(a)	(b)	(c)	(d)	(e)	Final score
ResNet50	87.29	81.90	84.96	88.56	67.73	57.58
ResNet101	87.73	83.40	86.32	91.07	74.42	63.06
DenseNet121	88.87	84.37	85.41	89.94	72.83	62.71
EfficientNet-B0	89.71	85.59	84.24	86.42	67.06	59.63
EfficientNet-B2	90.69	86.95	86.66	91.75	69.30	64.66
Combined 0	90.69	86.95	86.66	91.75	74.42	66.48

2.2 Metadata Only Models

Table 3. BACC scores of the hierarchical classifier using metadata only.

MLP	(a)	(b)	(c)	(d)	(e)	Final score
BACC	74.55	71.85	59.75	59.73	49.93	27.83

3 Combination of images with metadata

3.1 Concatenation

Table 4. BACC scores using the concatenation method on the flat and hierarchical classifiers. Bold scores represent the best results of each classifier excluding the model Combined 1.

CNN\Classifier	(a)	(b)	(c)	(d)	(e)	Final score	Flat
ResNet50	89.65	83.79	89.40	91.00	72.52	65.88	78.46
ResNet101	89.48	87.12	88.88	92.31	77.76	69.09	79.05
DenseNet121	92.21	85.92	87.44	91.63	75.86	68.26	75.91
EfficientNet-B0	91.23	87.01	87.69	94.64	73.61	68.62	73.59
EfficientNet-B2	91.42	88.10	88.47	90.29	77.47	69.95	74.47
Combined 1	92.21	88.10	89.40	94.64	77.76	71.98	-

3.2 Multiplication

ReLU activation

Table 5. BACC scores using the multiplication method with a ReLU activation function on the flat and hierarchical classifiers. Bold scores represent the best results of each classifier excluding the model Combined 2.

CNN\Classifier	(a)	(b)	(c)	(d)	(e)	Final score	Flat
ResNet50	91.37	85.29	89.08	93.49	76.57	70.00	76.02
ResNet101	91.51	85.77	87.67	86.76	80.57	68.95	79.02
DenseNet121	91.37	87.55	90.07	88.62	75.36	69.96	74.53
EfficientNet-B0	91.20	87.43	88.74	90.13	73.45	65.70	70.91
EfficientNet-B2	91.47	87.75	88.34	94.53	76.13	69.92	73.32
Combined 2	91.94	87.75	90.54	95.05	80.57	73.99	-

3.3 Image feature reducer

Experiment 1: Number of metadata features, m

Table 7. BACC scores of the first experiment of the image feature reducer for the hierarchical and flat models. Each test is represented by the number of metadata neurons, m , and the number of image features, n . Bold scores represent the best results of each classifier.

	CNN	$100(m) + 233(n)$	$200(m) + 466(n)$	$250(m) + 583(n)$
(a)	ResNet101	89.17	90.28	90.20
	DenseNet121	91.51	91.84	91.18
	EfficientNet-B2	90.96	91.05	91.30
(b)	ResNet101	84.26	84.74	83.90
	DenseNet121	87.10	85.62	84.55
	EfficientNet-B2	86.01	88.00	87.16
(c)	ResNet101	88.65	89.98	88.34
	DenseNet121	87.99	88.06	87.20
	EfficientNet-B2	88.69	87.81	88.74
(d)	ResNet101	91.87	93.46	92.94
	DenseNet121	88.89		92.01
	EfficientNet-B2	92.49	92.23	93.24
(e)	ResNet101	75.09	75.62	76.94
	DenseNet121	73.65	72.80	74.71
	EfficientNet-B2	72.34	73.95	74.83
Final score	ResNet101	67.16	69.05	68.26
	DenseNet121	68.27	67.34	66.49
	EfficientNet-B2	68.85	69.65	70.16
Flat	ResNet101	77.40	76.66	77.45
	DenseNet121	73.70	74.90	74.93
	EfficientNet-B2	72.75	74.29	73.19

Table 8. BACC scores of the first experiment of the image feature reducer for the hierarchical and flat models. Each test is represented by the number of metadata neurons, d , and the correspondent number of image features, n . Bold scores represent the best results of each classifier.

	CNN	$100(d) + 233(n)$	$200(d) + 466(n)$	$250(d) + 583(n)$
(a)	ResNet50	89.50	89.23	90.32
	EfficientNet-B0	90.27	91.00	90.53
(b)	ResNet50	84.12	85.46	84.78
	EfficientNet-B0	86.88	86.24	86.17
(c)	ResNet50	85.72	87.55	87.66
	EfficientNet-B0	85.89	86.61	87.25
(d)	ResNet50	92.92	94.73	90.33
	EfficientNet-B0	93.01	92.67	90.72
(e)	ResNet50	77.52	75.58	72.12
	EfficientNet-B0	71.64	71.03	71.87
Final score	ResNet50	65.11	67.26	64.37
	EfficientNet-B0	63.06	65.23	65.94
Flat	ResNet50	73.65	74.56	74.34
	EfficientNet-B0	71.65	71.60	73.02

Experiment 2: Ratio of image features, r

Table 9. BACC scores of the second experiment of the image feature reducer for the hierarchical and flat models. Each test is represented by the its ratio (r), the number of metadata neurons, d , and the correspondent number of image features, n . Bold scores represent the best results of each classifier.

	CNN	$r=0.5$ $200(d)+200(n)$	$r=0.6$ $200(d)+300(n)$	$r=0.7$ $200(d)+466(n)$	$r=0.8$ $200(d)+800(n)$
(a)	DenseNet121	90.69	91.70	91.84	90.81
	EfficientNet-B0	90.56	90.76	91.00	90.67
(b)	DenseNet121	86.36	84.39	85.62	84.74
	EfficientNet-B0	86.89	86.74	86.24	86.58
(c)	DenseNet121	88.44	88.25	88.06	88.29
	EfficientNet-B0	88.27	88.22	86.61	88.69
(d)	DenseNet121	89.08	93.33	90.37	92.35
	EfficientNet-B0	92.06	94.04	92.67	92.89
(e)	DenseNet121	75.72	75.83	72.80	76.24
	EfficientNet-B0	69.89	72.42	71.03	70.49
Final score	DenseNet121	67.11	69.94	67.34	67.72
	EfficientNet-B0	64.35	66.20	65.23	63.44
Flat	DenseNet121	73.34	74.81	74.90	75.43
	EfficientNet-B0	71.43	72.65	71.60	72.47

Table 10. BACC scores of the second experiment of the image feature reducer for the hierarchical and flat models. Each test is represented by the its ratio (r), the number of metadata neurons, m , and the correspondent number of image features, n . Bold scores represent the best results of each classifier.

	CNN	$r=0.5$ $200(m)+200(n)$	$r=0.6$ $200(m)+300(n)$	$r=0.7$ $200(m)+466(n)$	$r=0.8$ $200(m)+800(n)$
(a)	ResNet50	89.90	89.89	89.32	88.95
	ResNet101	91.07	90.70	90.28	90.10
	EfficientNet-B2	91.33	91.71	91.05	91.60
(b)	ResNet50	88.55	83.58	85.46	83.77
	ResNet101	86.68	84.22	84.74	84.92
	EfficientNet-B2	87.66	87.60	88.00	87.32
(c)	ResNet50	88.65	89.61	87.55	86.02
	ResNet101	88.70	90.12	89.98	86.88
	EfficientNet-B2	88.27	88.22	87.81	88.69
(d)	ResNet50	91.85	96.12	94.73	93.30
	ResNet101	94.40	91.66	93.46	95.95
	EfficientNet-B2	94.66	93.24	92.23	92.21
(e)	ResNet50	73.05	76.34	75.58	71.89
	ResNet101	77.66	78.96	75.62	76.62
	EfficientNet-B2	74.53	74.01	73.95	74.34
Final score	ResNet50	66.54	69.41	67.26	62.14
	ResNet101	72.70	71.77	69.05	68.29
	EfficientNet-B2	70.24	71.09	69.65	69.84
Flat	ResNet50	76.05	76.16	74.56	74.44
	ResNet101	72.87	77.76	76.66	79.08
	EfficientNet-B2	73.30	74.42	74.29	74.29

3.4 Comparison between models

Table 11. BACC scores of all the models best performance. Cells highlighted in green represent the best score of each classifier. The highlights of the individual classifiers, (a)-(e), also represent the classifiers of the Combined 4 model. Note that the flat score is from the method the Combined model belongs to.

Model\Classifier	(a)	(b)	(c)	(d)	(e)	Final score	Flat
Only Metadata	74.55	71.85	59.75	59.73	49.93	27.83	35.30
Only Images - Combined 0	90.69	86.95	86.66	91.75	69.30	66.48	73.05
Concatenation - Combined 1	92.21	88.10	89.40	94.64	77.76	71.98	79.05
Multiplication - Combined 2	91.94	87.75	90.54	95.05	80.57	73.99	79.02
Image feature reducer - Combined 3	91.84	88.00	90.12	96.12	78.96	72.92	79.08
Combined 4	92.21	88.10	90.54	96.12	80.57	73.82	-

Table 12. Mistakes of each individual classifier of the Combined 2 and Combined 4 model.

Model/Classifier	(a) (mel,n-mel)	(b) (MEL,NV)	(c) (ben,mal)	(d) (BKL,DF,VASC)	(e) (AK,BCC,SCC)	Total number of mistakes
Combined 2	387 (148,239)	311 (156,155)	117 (54,63)	12 (7,3,2)	123 (43,58,22)	950
Combined 4	355 (166,189)	350 (125,225)	117 (54,63)	15 (11,3,1)	120 (41,57,22)	965