Supplementary material: comparison of deep transfer learning strategies for digital pathology

Romain Mormont Pierre Geurts Raphaël Marée
University of Liège, Belgium
{r.mormont, p.geurts, raphael.maree}@uliege.be

A. Review about used networks

Architecture	Publications
Custom	[10, 16]
AlexNet (and variants)	[14, 2, 1, 13, 15, 11, 8]
GoogLeNet (InceptionV1)	[14, 2]
VGG16	[7, 2, 1, 16, 6, 11]
InceptionV3	[7, 4, 5]
ResNet50 (and variants)	[16]

Table 1. Networks used in recent publications for deep transfer learning in medical imaging.

B. Selected hyperparameters for the classifiers and the baseline

When we train a classifier on extracted features, we tune its hyperparameters by cross-validation. For SVM, we tune the penalty parameter C with values taken in $\left\{10^{-10}, 10^{-9}, ..., 10\right\}$. For extremely randomized trees, we grow fully expanded trees and tune the number of features evaluated at each split k among $\left\{1, \sqrt{n_f}, \frac{n_f}{2}, n_f\right\}$ where n_f is the total number of features. For the single layer perceptron, we tune the number of iterations among $\{1000, 2500, 5000, 10000\}$ and the learning rate among $\left\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\right\}$. All the other parameters values are the ones provided by default by the scikit-learn package [12]. More precisely, we use: the Adam [9] optimizer with default parameters, an L2 penalty on the weights with α set to 10^{-4} and a batch size of 200 samples.

For the baseline ET-FL, we tune the size of the windows (w_{min}, w_{max}) among all valid size ranges with `min_size` in $\{0.0, 0.25, 0.5, 0.75\}$ and `max_size` in $\{0.25, 0.5, 0.75, 1.0\}$ and the colorspace L among $\{\texttt{TRGB}, \texttt{HSV}\}$ where `TRGB` and `HSV` are respectively the normalized RGB and hue-saturation-value colorspaces. The number of extracted subwindow per image w is taken such that the total number of subwindows for the dataset w_t is approximately 1 million. The other fixed parameters are T, the number of trees, l_{min}, the minimum number of samples in a leave of a tree. The value for those parameters as well as the selected values for tuned parameters are given in Table 2.

C. Best features

In order to investigate the best features (according to feature importances) for a given network, we compute the minimum size of the best features subset so that at least one feature is in this subset for all datasets. The resulting sizes for all networks are given in Table 3. We also compute the percentages of overlap between subsets of selected features by RFE for all pairs of datasets (see Table 10)

D. Features and cut points information

In the paper, we investigate the usage of features from inside the networks. Therefore, we have selected the layers from which we would extract the features. As stated in the paper, we limit the number of possible cut points to the bottlenecks of the networks, a bottleneck being a point were several paths are merged into a single one. For ResNet, we have selected the ReLU activations of the merging layer after each residual block. For IncResV2, considering all the bottlenecks after inception-resnet and reduction blocks yields approximately fifty possible cut points, that is more than 100k features. We have decided to subsample those cut points to obtain a number of features closer to those of other networks while covering the network as uniformly as possible along its depth. As far as DenseNet is concerned, we extract features only at the end of dense blocks and after pooling blocks. Indeed, extracting features inside dense blocks would have resulted in duplicated features.

In Table 4 are given the information about the generated features vectors for the "*Last layer*", "*Merging features across networks*" and "*Merging layers features*" experiments. Information about the dimensions of the extracted features from inside the networks (before global average pooling) for ResNet, IncResV2 and DenseNet are respectively given in Tables 5, 6 and 7. The layer names given in those tables are the the ones given by the Keras package

Datasets	Fixed						Tuned (best)		
	T	w	w_t	l_{min}	k	C	w_{min}	w_{max}	L
C	20	551	1000616	1000	384	0.01	0.25	0.50	TRGB
G	20	69	1007745	1000	384	0.01	0.00	0.75	HSV
P	20	743	1000078	1000	384	0.01	0.25	0.50	HSV
N	20	1265	1000615	1000	384	0.01	0.00	0.25	HSV
B	20	55	1004355	1000	384	0.01	0.00	0.75	HSV
M	20	261	1001457	1000	384	0.01	0.25	0.75	TRGB
L	20	184	1001512	1000	384	0.01	0.25	0.50	HSV
H	20	228	1002516	1000	384	0.01	0.25	0.50	TRGB

Table 2. Hyperparameters for ET-FL.

$\mathcal{N}$	# feat.	% feat.
Mobile	753	73.54
IncResV2	1277	83.14
IncV3	1537	75.05
ResNet	1803	88.04
VGG16	409	79.88
VGG19	392	76.56
DenseNet	1477	76.93

Table 3. Given a network, this table gives the best features subset minimum size so that there is at least one feature that is in this subset for all datasets.

$\mathcal{N}$	Last layer	Merged layers	
	# feat.	# feat.	# cut
Mobile	1024	/	/
DenseNet	1920	7744	9
IncResV2	1536	17088	12
ResNet	2048	15168	17
IncV3	2048	/	/
VGG19	512	/	/
VGG16	512	/	/
Total	9600	/	/

Table 4. Number of features extracted for the "*Last layer*" experiment. Total number of features for the "*Merging features across networks*". Number of features and cut points for the "*Mering layers features*" experiment.

[3].

E. Charts for "merging layers" experiment

The extracted average importances per layer for the "*Merging layers features*" experiment are given in Figure 1.

F. Features selected with RFE

A summary of the number of selected features and the cross-validation curves for all datasets and networks are respectively given in Table 8 and Figure 2.

Layer l (name)	Feat. maps dim.		
	h_a	w_a	d
activation_1	112	112	64
activation_4	55	55	256
activation_7	55	55	256
activation_10	55	55	256
activation_13	28	28	512
activation_16	28	28	512
activation_19	28	28	512
activation_22	28	28	512
activation_25	14	14	1024
activation_28	14	14	1024
activation_31	14	14	1024
activation_34	14	14	1024
activation_37	14	14	1024
activation_40	14	14	1024
activation_43	7	7	2048
activation_46	7	7	2048
activation_49 (last)	7	7	2048
Total	/	/	**15168**

Table 5. Name and dimensions of the layers extracted from inside ResNet for the "*Merging features across layers*" and "*Inner layers*" experiments.

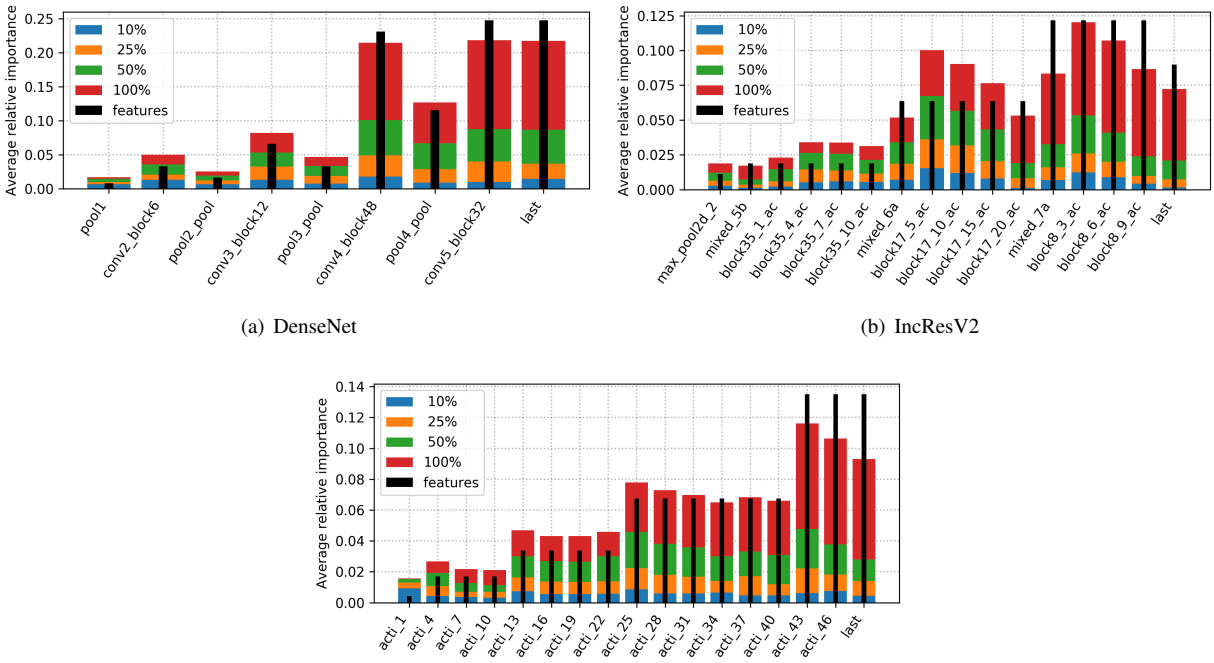
G. Detailed scores

Detailed scores for all datasets, experiments and networks are given in Tables 11 and 12.

H. Acknowledgments

We thank our collaborators for bringing images and annotations.

- *CellInclusion and ProliferativePatterns*: Caroline Degand and Isabelle Salmon (Erasme Hospital, Université Libre de Bruxelles)
- *Breast*: Michel Reginster and Philippe Delvenne (University Hospital, Liège)
- *Necrosis*: Natacha Leroi and Philippe Martinive (GIGA-Cancer, ULiege)



(a) DenseNet

(b) IncResV2

(c) ResNet

Figure 1. Average relative importances (across datasets) brought by each extracted layer of a network. The black bars quantify the proportion of features of each layer. The colors indicate the information brought by features of decreasing importances: blue and red features are respectively the most informative and least informative ones. Blue, orange, green and red bars regroup importances of features that respectively and cumulatively bring 10%, 25%, 50% and 100% of the information for predicting the outcome.

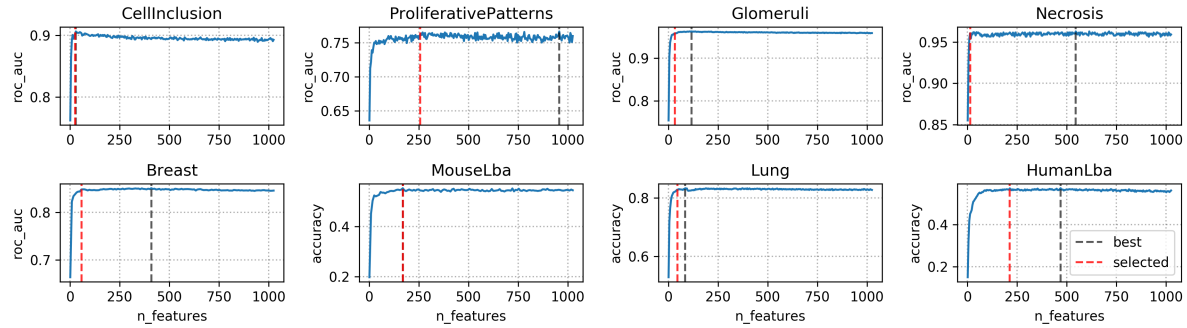
Layer l (name)	**Feat. maps dim.**		
	h_a	w_a	d
max_pooling2d_2	25	25	192
mixed_5b	25	25	320
block35_1_ac	25	25	320
block35_4_ac	25	25	320
block35_7_ac	25	25	320
block35_10_ac	25	25	320
mixed_6a	12	12	1088
block17_5_ac	12	12	1088
block17_10_ac	12	12	1088
block17_15_ac	12	12	1088
block17_20_ac	12	12	1088
mixed_7a	5	5	2080
block8_3_ac	5	5	2080
block8_6_ac	5	5	2080
block8_9_ac	5	5	2080
conv_7b_ac (last)	5	5	1536
Total	/	/	**17088**

Table 6. Name and dimensions of the layers extracted from inside IncResV2 for the "*Merging features across layers*" and "*Inner layers*" experiments.

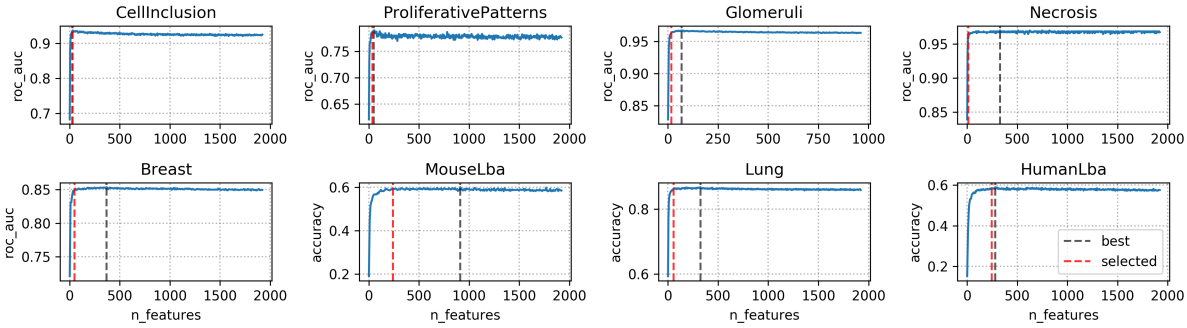
Layer l (name)	**Feat. maps dim.**		
	h_a	w_a	d
pool1	56	56	64
conv2_block6_concat	56	56	256
pool2_pool	28	28	128
conv3_block12_concat	28	28	512
pool3_pool	14	14	256
conv4_block48_concat	14	14	1792
pool4_pool	7	7	896
conv5_block32_concat	7	7	1920
bn (last)	7	7	1920
Total	/	/	**7744**

Table 7. Name and dimensions of the layers extracted from inside DenseNet for the "*Merging features across layers*" and "*Inner layers*" experiments.

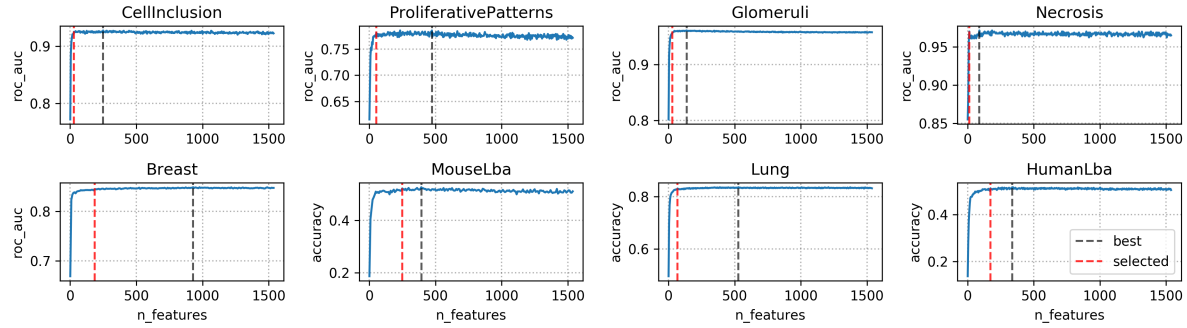
- *HumanLba*: Sandrine Rorive and Isabelle Salmon (Erasme Hospital, Université Libre de Bruxelles)
- *MouseLba*: Natacha Rocks, Christine Fink, Fabienne Perin, and Didier Cataldo (GIGA-Cancer, ULiege)
- *Lung*: Natacha Rocks, Christine Fink, Fabienne Perin, and Didier Cataldo (GIGA-Cancer, ULiege)
- *Glomeruli*: Vannary Meas-Yedid and Jean-Christophe Olivo-Marin (Pasteur Institute Paris), and Eric Thervet



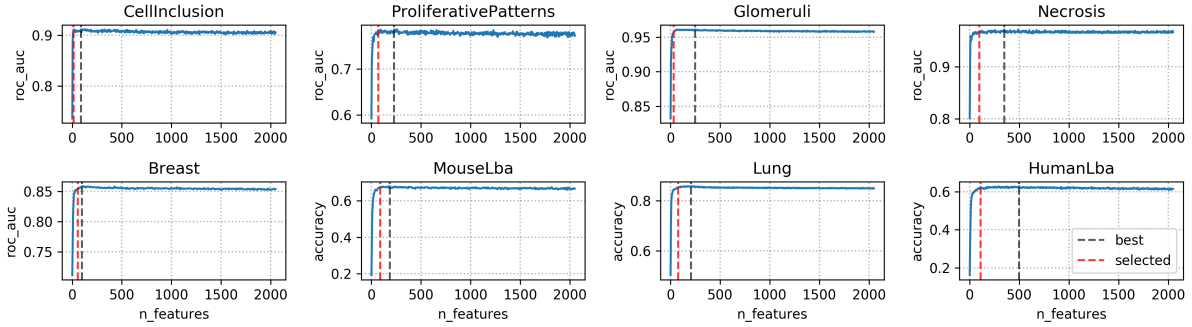
(a) Mobile



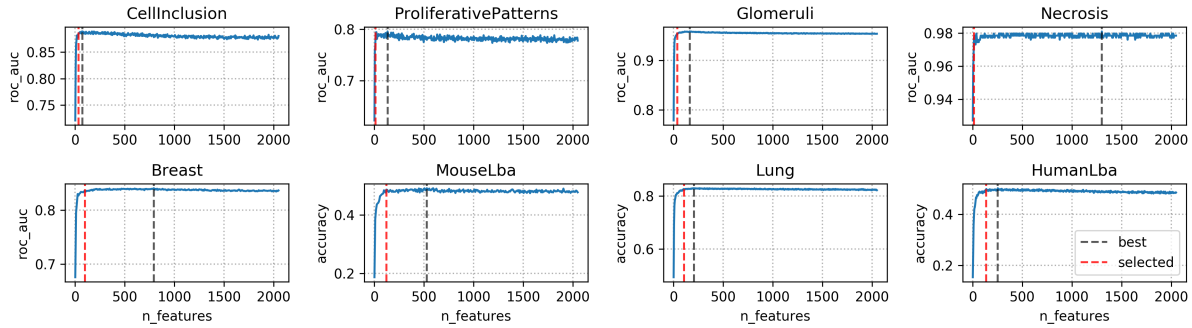
(b) DenseNet



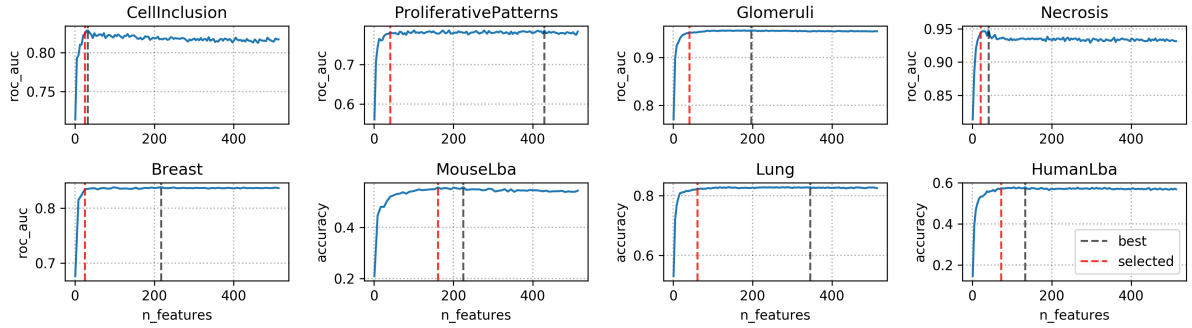
(c) IncResV2



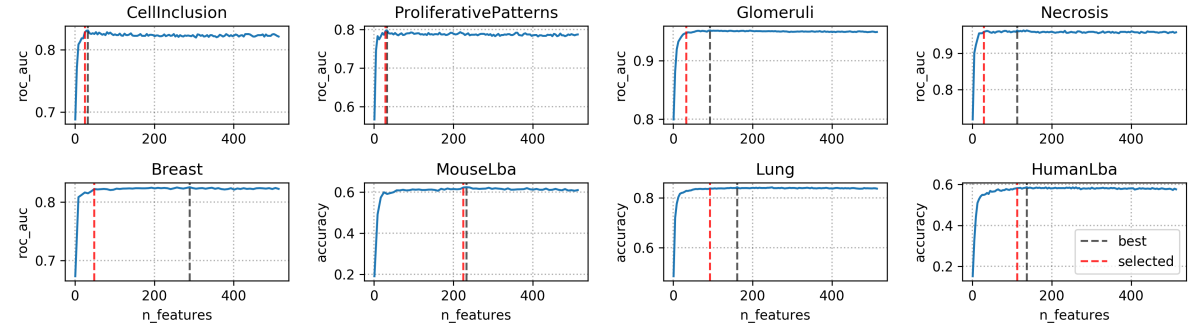
(d) ResNet



(e) IncV3



(f) VGG19



(g) VGG16

Figure 2. Cross validation curves from recursive feature elimination (RFE).

$\mathcal{N}$	**Number of features**							
	C	P	G	N	B	M	L	H
Mobile	25	257	33	13	57	169	45	213
DenseNet	25	37	17	13	49	241	57	245
IncResV2	29	53	29	13	185	249	69	173
ResNet	13	69	33	97	57	89	77	109
IncV3	33	13	37	13	97	121	105	137
VGG19	25	41	41	21	25	161	61	73
VGG16	25	29	33	29	49	225	93	113

Table 8. Number of features selected by RFE for all datasets and networks.

(Georges Pompidou European Hospital Paris)

$\mathcal{N}$	**Proportion of features (%)**								Average
	C	P	G	N	B	M	L	H	
Mobile	2.44	25.10	3.22	1.27	5.57	16.50	4.39	20.80	9.91
DenseNet	1.30	1.93	0.89	0.68	2.55	12.55	2.97	12.76	4.45
IncResV2	1.89	3.45	1.89	0.85	12.04	16.21	4.49	11.26	6.51
ResNet	0.63	3.37	1.61	4.74	2.78	4.35	3.76	5.32	3.32
IncV3	1.61	0.63	1.81	0.63	4.74	5.91	5.13	6.69	3.39
VGG19	4.88	8.01	8.01	4.10	4.88	31.45	11.91	14.26	10.94
VGG16	4.88	5.66	6.45	5.66	9.57	43.95	18.16	22.07	14.55
Average	2.52	6.88	3.41	2.56	6.02	18.70	7.26	13.31	**7.58**

Table 9. Proportion of features selected by RFE for all datasets and networks.

(a) Mobile

Dataset	C	P	G	N	B	M	L	H
C		1	3	0	0	3	4	2
P	20		27	53	40	33	33	34
G	4	3		7	14	4	15	6
N	0	2	3		5	1	4	0
B	0	8	24	23		5	17	8
M	24	22	21	15	17		26	31
L	8	5	21	15	14	7		6
H	20	28	39	15	33	40	31	

(b) IncResV2

Dataset	C	P	G	N	B	M	L	H
C		1	0	0	3	2	7	3
P	3		3	0	6	4	2	5
G	0	1		0	1	2	0	1
N	0	0	0		0	0	0	0
B	20	22	10	7		14	20	13
M	17	22	17	15	18		20	21
L	17	3	0	0	7	5		7
H	20	17	10	0	12	14	18	

(c) IncV3

Dataset	C	P	G	N	B	M	L	H
C		0	2	0	4	2	3	3
P	0		0	7	0	1	2	1
G	3	0		0	8	1	6	0
N	0	7	0		2	2	1	1
B	12	0	21	15		10	13	10
M	9	15	5	23	13		21	20
L	12	23	18	15	14	19		10
H	15	15	2	15	15	23	14	

(d) ResNet

Dataset	C	P	G	N	B	M	L	H
C		2	0	2	3	3	0	2
P	15		0	7	7	7	14	11
G	0	0		2	5	2	5	1
N	15	10	6		19	11	11	9
B	15	5	9	11		6	16	11
M	23	10	6	10	10		10	23
L	0	15	12	9	22	9		10
H	23	18	6	10	21	29	14	

(e) VGG16

Dataset	C	P	G	N	B	M	L	H
C		10	3	10	8	6	5	7
P	12		9	13	14	9	14	9
G	4	10		13	12	8	17	6
N	12	13	12		16	7	8	11
B	16	24	18	27		16	18	19
M	60	72	60	58	73		62	78
L	20	44	48	27	34	25		26
H	32	37	21	44	44	39	32	

(f) VGG19

Dataset	C	P	G	N	B	M	L	H
C		4	2	4	12	6	8	8
P	8		14	19	12	11	14	16
G	4	14		0	24	10	21	6
N	4	9	0		20	6	9	8
B	12	7	14	23		7	13	9
M	44	46	41	52	48		50	45
L	20	22	31	28	32	19		16
H	24	29	12	28	28	20	19	

(g) DenseNet

Dataset	C	P	G	N	B	M	L	H
C		2	0	0	0	1	3	2
P	4		11	0	2	3	3	3
G	0	5		0	6	0	1	0
N	0	0	0		0	1	0	0
B	0	2	17	0		3	5	2
M	16	21	11	23	16		21	24
L	8	5	5	0	6	5		7
H	28	21	11	15	12	24	31	

Table 10. Percentages of overlap between features selected by RFE on the studied datasets. The tables can be read as follows: the number at row i and column j is the percentage of features among the ones selected for dataset j that were also selected for the dataset i.

Experiment	$\mathcal{C}$	$\mathcal{N}$	**C**	**P**	**G**	**N**	**B**	**M**	**L**	**H**
Baseline	ET-FL		0.9250	0.8268	0.9551	0.9805	0.9345	0.7568	0.8547	0.6960
Last layer	SVM	Mobile	0.9749	0.8844	0.9935	0.9953	0.9427	0.7611	0.9043	0.6882
		DenseNet	0.9794	0.8852	0.9938	0.9864	0.9257	0.7010	0.9133	0.7820
		IncResV2	0.9795	0.8698	0.9928	0.9982	0.9485	0.6566	0.9077	0.7351
		ResNet	0.9748	0.8893	0.9924	0.9882	0.9372	0.7633	0.9122	0.7791
		IncV3	0.9722	0.8670	0.9910	0.9964	0.8951	0.6371	0.9088	0.7175
		VGG19	0.8853	0.8654	0.9860	0.9905	0.9241	0.7237	0.8885	0.7302
		VGG16	0.8824	0.8808	0.9859	0.9893	0.9413	0.7438	0.9020	0.7028
Last layer	ET	Mobile	0.9608	0.8848	0.9854	0.9840	0.9487	0.5872	0.8648	0.7126
		DenseNet	0.9726	0.8889	0.9891	0.9870	0.9556	0.6381	0.8874	0.7410
		IncResV2	0.9618	0.8699	0.9824	0.9953	0.9408	0.4789	0.8570	0.6676
		ResNet	0.9634	0.8832	0.9758	0.9929	0.9507	0.6186	0.8851	0.7752
		IncV3	0.9481	0.8795	0.9793	0.9929	0.9428	0.4583	0.8300	0.6305
		VGG19	0.8551	0.8430	0.9795	0.9861	0.9231	0.5379	0.8536	0.6989
		VGG16	0.8412	0.8791	0.9690	0.9888	0.9254	0.5791	0.8659	0.6833
Last layer	FC	Mobile	0.9796	0.8661	0.9794	0.9935	0.9603	0.7941	0.8986	0.6823
		DenseNet	0.9822	0.8668	0.8316	0.9852	0.9482	0.7291	0.9054	0.7664
		IncResV2	0.9756	0.8676	0.9729	0.9976	0.9597	0.6598	0.9043	0.7038
		ResNet	0.9726	0.8670	0.9771	0.9899	0.9583	0.7996	0.9133	0.7674
		IncV3	0.9714	0.8417	0.9796	0.9893	0.9377	0.6538	0.8998	0.7038
		VGG19	0.8447	0.8553	0.9661	0.9899	0.9237	0.6636	0.8863	0.7410
		VGG16	0.8298	0.8718	0.9573	0.9852	0.9421	0.6956	0.9088	0.7185
Feature selection	SVM	Mobile	0.9610	0.7876	0.9794	0.9870	0.9597	0.7421	0.8682	0.6618
		DenseNet	0.9347	0.8212	0.8316	0.9888	0.9436	0.6614	0.7984	0.6931
		IncResV2	0.9665	0.8476	0.9729	0.9976	0.9443	0.6403	0.8682	0.7214
		ResNet	0.9578	0.8337	0.9771	0.9722	0.9492	0.7438	0.8806	0.7644
		IncV3	0.9562	0.8308	0.9796	0.9964	0.9436	0.6430	0.8750	0.6843
		VGG19	0.8284	0.8488	0.9661	0.9888	0.8860	0.6750	0.8784	0.7038
		VGG16	0.8071	0.8810	0.9573	0.9899	0.9131	0.7362	0.8941	0.7038
Feature selection	ET	Mobile	0.9617	0.8798	0.9799	0.9888	0.9581	0.6582	0.8694	0.7400
		DenseNet	0.9676	0.8861	0.9843	0.9994	0.9489	0.6939	0.8919	0.6667
		IncResV2	0.9609	0.8646	0.9743	0.9944	0.9421	0.5330	0.8491	0.6500
		ResNet	0.9503	0.8786	0.9799	0.9852	0.9488	0.6961	0.8863	0.7038
		IncV3	0.9473	0.8410	0.9786	0.9959	0.9466	0.5531	0.8378	0.7703
		VGG19	0.8506	0.8492	0.9732	0.9947	0.9186	0.5850	0.8468	0.7019
		VGG16	0.8232	0.8774	0.9659	0.9953	0.9282	0.6349	0.8615	0.6745
Merging networks	ET	merged	0.9897	0.8573	0.9948	0.9858	0.8851	0.8169	0.9155	0.7928
	SVM	merged	0.9784	0.8984	0.9912	0.9864	0.9549	0.6896	0.8615	0.6063
Merging layers	SVM	DenseNet	0.9757	0.8090	0.9835	0.9870	0.9470	0.7042	0.8840	0.7761
		IncResV2	0.9808	0.8418	0.9920	0.9964	0.9559	0.7031	0.9155	0.7761
		ResNet	0.9789	0.8576	0.9927	0.9953	0.9234	0.7941	0.9268	0.7977
Merging layers	ET	DenseNet	0.9605	0.8892	0.9875	0.9911	0.9588	0.6993	0.8818	0.7370
		IncResV2	0.9799	0.8906	0.9944	0.9920	0.9639	0.6495	0.8897	0.7370
		ResNet	0.9424	0.8787	0.9847	0.9929	0.9619	0.7080	0.8885	0.7683
Fine-tuning	SVM	DenseNet	0.9883	0.8556	0.9944	0.9870	0.9777	0.8342	0.9119	0.8553
		IncResV2	0.9841	0.8377	0.9909	0.9941	0.9403	0.6847	0.9039	0.7390
		ResNet	0.9921	0.8705	0.9897	0.9941	0.9637	0.8147	0.9119	0.8456
Fine-tuning	ET	DenseNet	0.9828	0.8965	0.9950	0.9876	0.9827	0.7887	0.8982	0.8094
		IncResV2	0.9769	0.8776	0.9850	0.9929	0.9477	0.5406	0.8446	0.7048
		ResNet	0.9909	0.8806	0.9879	0.9870	0.9772	0.7763	0.8845	0.8289
Fine-tuning	Net	DenseNet	0.9892	0.8797	0.9977	0.9893	0.9835	0.8483	0.9405	0.8641
		IncResV2	0.9851	0.8795	0.9971	0.9929	0.9873	0.8727	0.9165	0.8182
		ResNet	0.9926	0.8778	0.9953	0.9970	0.9827	0.8288	0.8971	0.8416
Metric			ROC AUC					Accuracy (multi-class)		

Table 11. Detailed scores for all datasets and for the "*Last layer*", "*Feature selection*", "*Merging features across networks*", "*Merging features across layers*" and "*Fine-tuning*" experiments.

$\mathcal{N}$	**Layer** l	**C**	**P**	**G**	**N**	**B**	**M**	**L**	**H**
Baseline / ET-FL		0.9250	0.8268	0.9551	0.9805	0.9345	0.7568	0.8547	0.6960
ResNet	activation_1	0.7720	0.8415	0.9275	0.9811	0.9100	0.4946	0.8153	0.5758
	activation_4	0.8283	0.8275	0.9723	0.9964	0.9390	0.7308	0.8806	0.6285
	activation_7	0.8456	0.8276	0.9772	0.9888	0.9351	0.7275	0.8930	0.6755
	activation_10	0.8574	0.8292	0.9759	0.9882	0.9439	0.6717	0.8930	0.6491
	activation_13	0.8859	0.8608	0.9824	0.9888	0.9483	0.7313	0.9077	0.6188
	activation_16	0.8975	0.8418	0.9860	0.9876	0.9478	0.7356	0.9054	0.6598
	activation_19	0.8877	0.8503	0.9892	0.9888	0.9499	0.7270	0.9077	0.6510
	activation_22	0.9244	0.8763	0.9892	0.9882	0.9555	0.7010	0.9223	0.6940
	activation_25	0.9506	0.8785	0.9933	0.9858	0.9639	0.7736	0.9223	0.7253
	activation_28	0.9489	0.8884	0.9935	0.9876	0.9638	0.8131	0.9245	0.7634
	activation_31	0.9519	0.8724	0.9938	0.9876	0.9659	0.7996	0.9201	0.7351
	activation_34	0.9584	0.8947	0.9940	0.9876	0.9606	0.7514	0.9223	0.7977
	activation_37	0.9671	0.8959	0.9942	0.9876	0.9663	0.7600	0.9280	0.7996
	activation_40	0.9621	0.8894	0.9949	0.9864	0.9664	0.7914	0.9155	0.8113
	activation_43	0.9710	0.8950	0.9942	0.9852	0.9648	0.8017	0.9223	0.8074
	activation_46	0.9712	0.8848	0.9937	0.9870	0.9652	0.7860	0.9291	0.8094
	activation_49 (last)	0.9748	0.8893	0.9924	0.9882	0.9640	0.7860	0.9122	0.7791
DenseNet	pool1	0.7187	0.8276	0.8994	0.9533	0.9227	0.4821	0.7826	0.4653
	conv2_block6_concat	0.7982	0.8374	0.9609	0.9905	0.9374	0.6300	0.8536	0.5259
	pool2_pool	0.8185	0.8296	0.9570	0.9893	0.9510	0.6235	0.8570	0.5337
	conv3_block12_concat	0.9024	0.8361	0.9861	0.9882	0.9522	0.6696	0.9020	0.6823
	pool3_pool	0.9309	0.8900	0.9832	0.9893	0.9382	0.6300	0.9088	0.6686
	conv4_block48_concat	0.9803	0.8876	0.9962	0.9870	0.9699	0.8012	0.9223	0.7674
	pool4_pool	0.9843	0.8984	0.9954	0.9870	0.9613	0.7703	0.9268	0.7859
	conv5_block32_concat	0.9862	0.8981	0.9955	0.9917	0.9623	0.7806	0.9201	0.7879
	bn (last)	0.9784	0.8867	0.9931	0.9852	0.9538	0.7573	0.9043	0.7967
IncResV2	max_pooling2d_2	0.8403	0.8091	0.9716	0.9941	0.9340	0.6143	0.8851	0.6158
	mixed_5b	0.8265	0.8146	0.9771	0.9905	0.9424	0.6945	0.8897	0.6461
	block35_1_ac	0.8325	0.8412	0.9776	0.9941	0.9412	0.6576	0.8897	0.6373
	block35_4_ac	0.8673	0.8770	0.9834	0.9923	0.9556	0.6495	0.8998	0.6716
	block35_7_ac	0.8981	0.8709	0.9844	0.9935	0.9590	0.6354	0.9043	0.7048
	block35_10_ac	0.9219	0.8692	0.9900	0.9935	0.9616	0.6549	0.9110	0.7253
	mixed_6a	0.9445	0.8747	0.9920	0.9953	0.9706	0.7172	0.9088	0.7439
	block17_5_ac	0.9681	0.8665	0.9945	0.9917	0.9695	0.8066	0.9190	0.7713
	block17_10_ac	0.9711	0.8687	0.9958	0.9935	0.9720	0.8137	0.9234	0.7674
	block17_15_ac	0.9762	0.8939	0.9960	0.9923	0.9622	0.7985	0.9144	0.7419
	block17_20_ac	0.9860	0.8948	0.9957	0.9923	0.9649	0.7741	0.9155	0.7693
	block8_3_ac	0.9873	0.8905	0.9959	0.9953	0.9676	0.7790	0.9190	0.7273
	block8_6_ac	0.9868	0.8871	0.9953	0.9923	0.9686	0.7562	0.9212	0.7468
	block8_9_ac	0.9824	0.8773	0.9946	0.9964	0.9632	0.7427	0.9144	0.7468
	mixed_7a	0.9868	0.8934	0.9962	0.9959	0.9602	0.7893	0.9178	0.7214
	conv_7b_ac (last)	0.9773	0.8615	0.9926	0.9982	0.9619	0.6766	0.8998	0.7361
Metric		ROC AUC					Accuracy (multi-class)		

Table 12. Detailed scores for all datasets and for the "*Inner layer*" experiment.

References

[1] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 1195–1200. IEEE, 2016.

[2] N. Bayramoglu and J. Heikkilä. Transfer learning for cell nuclei classification in histopathology images. In *European Conference on Computer Vision*, pages 532–539. Springer, 2016.

[3] F. Chollet et al. Keras. `https://github.com/keras-team/keras`, 2015.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

[5] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[6] L. Hou, K. Singh, D. Samaras, T. M. Kurc, Y. Gao, R. J. Seidman, and J. H. Saltz. Automatic histopathology image analysis with cnns. In *Scientific Data Summit (NYSDS), 2016 New York*, pages 1–6. IEEE, 2016.

[7] B. Kieffer, M. Babaie, S. Kalra, and H. Tizhoosh. Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. *arXiv preprint arXiv:1710.05726*, 2017.

[8] E. Kim, M. Corte-Real, and Z. Baloch. A deep semantic mobile application for thyroid cytopathology. In *Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations*, volume 9789, page 97890A. International Society for Optics and Photonics, 2016.

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] O. Z. Kraus, B. T. Grys, J. Ba, Y. Chong, B. J. Frey, C. Boone, and B. J. Andrews. Automated analysis of high-content microscopy data with deep learning. *Molecular systems biology*, 13(4):924, 2017.

[11] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. Tizhoosh. A comparative study of cnn, bovw and lbp for classification of histopathological images. *arXiv preprint arXiv:1710.01249*, 2017.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[13] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, and V. Vaidya. Understanding the mechanisms of deep transfer learning for medical images. In *Deep Learning and Data Labeling for Medical Applications*, pages 188–196. Springer, 2016.

[14] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[15] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[16] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao. Deep transfer learning for modality classification of medical images. *Information*, 8(3):91, 2017.