Deep learning based Feature Selection Study

Sanghyuck Lee

2021. 12. 14.

To DO List

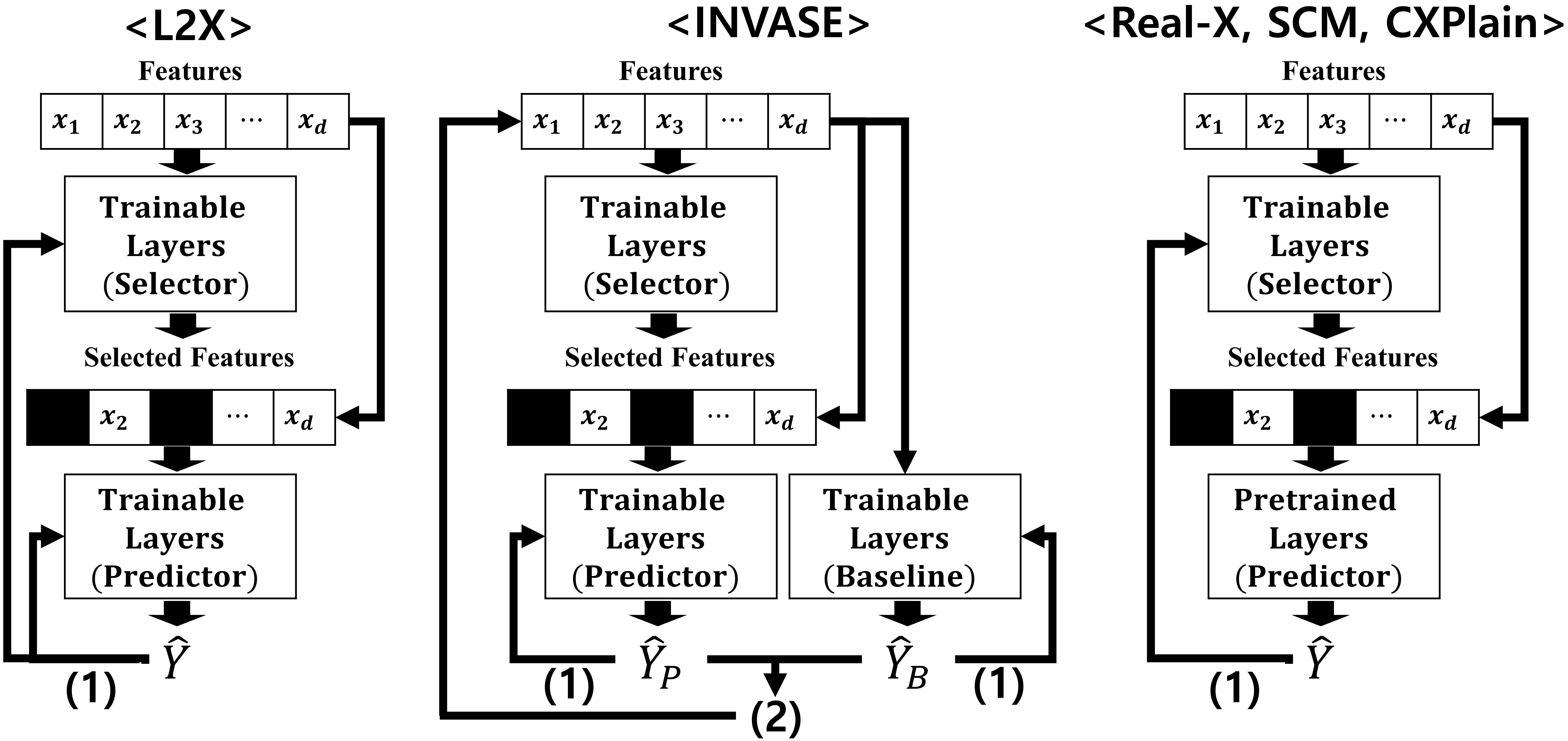
➢Last Week

- Study Related Works
- 갑상선 연구 보고 : INVASE 적용 파일럿 실험

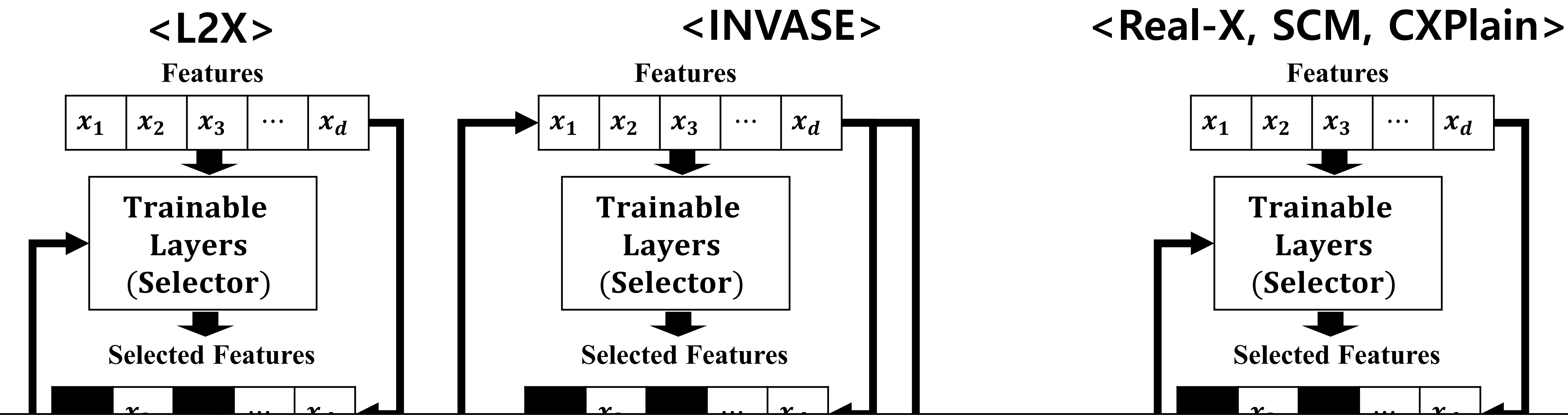
➢This Week

- Study Related Works
- 갑상선 연구 보고 : Activity 추론이 얼마나 어려운지 확인

L2X vs INVASE vs Real-X, SCM, CXPlain



L2X vs INVASE vs Real-X, SCM, CXPlain



Model	Dataset Modality
L2X	Synthetic, Text, Image (MNIST)
INVASE	Synthetic, Tabular
Real-X	Synthetic, Image (MNIST, Chest X-Rays)
SCM	Image (MNIST, FMNIST, CIFAR-10)
Cxplain	Image (MNIST, ImageNet)

Study Related Work

- ✓ L2X: (2018 AISTATS) Chen, Jianbo, et al. "Learning to explain: An information-theoretic perspective on model interpretation." International Conference on Machine Learning. PMLR, 2018.
- ✓ INVASE: (2018 ICLR) Yoon, Jinsung, James Jordon, and Mihaela van der Schaar. "INVASE: Instance-wise variable selection using neural networks." *International Conference on Learning Representations*. 2018.
- ✓ Cxplain: (NeurIPS 2019) Schwab, Patrick, and Walter Karlen. "Cxplain: Causal explanations for model interpretation under uncertainty." *arXiv preprint arXiv:1910.12336* (2019).
- ✓ Real-X: (2020 NeurIPS) Jethani, Neil, et al. "Can We Learn to Explain Chest X-Rays?: A Cardiomegaly Use Case."
- ✓ SCM: (2021 CVPR) Panda, Pranoy, Sai Srinivas Kancheti, and Vineeth N. Balasubramanian. "Instance-wise Causal Feature Selection for Model Interpretation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

75 Benchmark Image Datasets

To Find Complex Image Datasets, Ascending Datasets by Top SOTA model Accuracy

Popular Rank	Dataset	Accuracy
69	Red MiniImageNet 80% label noise	35.50
72	ImageNet-Sketch	39.10
68	Red MiniImageNet 60% label noise	45.10
67	Red MiniImageNet 40% label noise	48.06
66	Red MiniImageNet 20% label noise	51.42
75	Places365-Standard	60.30
34	**JFT-300M**	**60.62**
65	CIFAR-100, 60% Symmetric Noise	64.60
22	**Places205**	**66.80**
44	**Tiny-ImageNet**	**67.67**
59	ArtDL	70.25
56	ImageNet-10	72.90
9	**Clothing1M**	**75.19**
15	**DF20 - Mini**	**75.85**
51	CIFAR-100, 40% Symmetric Noise	75.91
47	DTD	76.60
19	**WebVision-1000**	**79.30**
10	**VTAB-1k**	**79.99**
23	**ObjectNet**	**80.00**
27	**Tiered ImageNet 5-way (5-shot)**	**80.15**
16	**DF20**	**80.45**
11	**mini WebVision 1.0**	**81.47**
35	Clothing1M (using clean data)	81.50
37	CelebA 64x64	82.00
60	PASCAL VOC 2007	83.00

Popular Rank	Dataset	Accuracy
24	iNaturalist	83.40
74	Chaoyang	83.40
17	ImageNet V2	84.00
30	Tiny ImageNet Classification	84.65
45	X-ray Images on MURA	84.72
31	ObjectNet (Bounding Box)	85.10
12	iNaturalist 2018	86.80
57	LabelMe	87.12
18	iNaturalist 2019	88.30
33	MAMe	88.95
38	Food-101	89.30
40	Food-101N	90.39
1	**ImageNet**	**90.88**
29	EMNIST-Balanced	91.05
7	**ImageNet ReaL**	**91.12**
50	CIFAR-10, 60% Symmetric Noise	91.30
36	CUB	92.95
64	Sports10	93.42
54	iCassava'19	93.68
25	CINIC-10	94.30
52	MultiMNIST	94.80
21	Oxford-IIIT Pets	94.90
58	Surrey ASL	94.90
49	CIFAR-10, 40% Symmetric Noise	95.37
55	cifar-10,4000	95.80

Popular Rank	Dataset	Accuracy
26	EMNIST-Letters	95.88
3	**Cifar-100**	**96.08**
14	Stanford Cars	96.32
73	CIFAR-10 (with noisy labels)	96.70
20	Fashion-MNIST	96.91
43	Noisy MNIST (Contrast)	97.25
48	SIPaKMeD	97.87
61	MNIST-rot-12	98.13
63	Herlev	98.32
41	Noisy MNIST (AWGN)	98.43
46	EuroSAT	98.65
4	**STL-10**	**98.66**
28	smallNORB	98.77
70	SARS-COV-2	98.93
5	**SVHN**	**99.10**
13	Kuzushiji-MNIST	99.15
42	Noisy MNIST (Motion)	99.20
62	MNIST-rot-12k (DA)	99.29
32	EMNIST-Digits	99.43
71	PlantVillage	99.48
2	**Cifar-10**	**99.50**
39	Imbalanced CUB-200-2011	99.60
53	QMNIST	99.67
8	**Flowers-102**	**99.76**
6	**MNIST**	**99.91**

Next Step

- Lightweight Deep Learning Model combining Feature Selection and Neural Architecture Search:

  1) Selector Experiments for Popular & Complex Image Datasets

     => Select some datasets

     => L2X, INVASE, Real-X, SCM, CXPlain

     + GradCAM, Saliency, SHAP, DeepSHAP, LIME,

  2) Lightweight Experiments with ProxylessNAS

     => Check Final Efficiency of Output Networks

갑상선 연구: 실험 세팅

- 데이터셋: Active / Inactive / Normal = 144 / **924** / 121
  Train / Validation / Test = 0.8 / 0.1 / 0.1
- 클래스 불균형 데이터셋
- Activity 추론이 얼마나 어려운지 확인해보기로 함.
- 실험 후보 모델 : AlexNet, DenseNet169, DenseNet264, EfficientNetB0, GoogLeNet, InceptionV2, MobileNetV3, ResNet50, ResNet101, ShuffleNet_v2_x1_0, ShuffleNet_v2_x2_0, VGGNet16, VGGNet19
- 실험 1 : 13개 모델과 각 11개 슬라이스에 대해서 비교 실험 (Adam)
- 실험 2 : 13개 모델과 1개 슬라이스에 대해서 Optimizer를 변경(Adam&SGD)해 비교 실험
- 실험 3 : 13개 모델과 11채널을 모두 입력해서 비교 실험 (Adam&SGD)
- 실험 4 : 불균형 데이터셋 딜링 스킬(Under-sampling, Over-sampling, Augmentation) 비교 실험

갑상선 연구

- 실험 1 : 13개 모델과 각 11개 슬라이스에 대해서 비교 실험
- AUC (one vs rest) : 39.24%~81.9%
- AUC (one vs one) : 41.82%~83.34%

순위	슬라이스	13개 모델 평균 AUC (one vs rest)
1	AX3	65.61
2	AX2	64.14
3	AX4	63.86
4	RT	63.03
5	AX1	62.85
6	AX5	62.15
7	LT	61.85
8	CO2	61.84
9	CO1	56.61
10	AX6	55.27
11	CO3	55.00

순위	모델	11개 슬라이스 평균 순위	11개 슬라이스 평균 AUC (one vs rest)
1	**DenseNet264**	**3**	**70.29**
2	**VGGNet19**	**3.64**	**68.48**
3	**VGGNet16**	**4.36**	**66.23**
4	**DenseNet169**	**4.45**	**68.04**
5	**AlexNet**	**4.73**	**66.39**
6	GoogLeNet	7.73	59.68
7	ShuffleNet_v2_x2_0	8	57.49
8	ResNet101	8.09	57.35
9	ResNet50	8.36	58.47
10	MobileNetV3	8.36	57.12
11	InceptionV2	8.55	58.26
12	ShuffleNet_v2_x1_0	10.64	54.58
13	EfficientNetB0	11.36	51.07

갑상선 연구

- 실험 2 : 13개 모델과 1개 슬라이스(AX3)에 대해서 Optimizer를 변경(Adam&SGD)해 비교 실험

순위	모델	Max(adam AUC, sgd AUC)	Adam	SGD	#Params (million)
1	DenseNet264	81.9	**81.9**	74.23	30
2	VGGNet19	80.48	76.96	**80.48**	139
3	VGGNet16	79.67	77.39	**79.67**	134
4	DenseNet169	73.48	**73.48**	73.09	12
5	**ResNet50**	72.26	61.19	**72.26**	23
6	AlexNet	71.97	**71.97**	69.26	57
7	**ShuffleNet_v2_x2_0**	70.46	54.97	**70.46**	19
8	**InceptionV2**	69.46	**69.46**	62.48	7
9	GoogLeNet	69	67.77	**69**	5
10	MobileNetV3	68.71	63.58	**68.71**	4
11	ResNet101	68.28	49.61	**68.28**	42
12	ShuffleNet_v2_x1_0	63.45	60.39	**63.45**	4
13	EfficientNetB0	61.98	44.25	**61.98**	4

갑상선 연구

- 실험 3 : 13개 모델과 11채널을 모두 입력해서 비교 실험 (Adam&SGD)

순위	모델	AX3만	11개 슬라이스	# Params
1	DenseNet264	**81.9**	75.53	30
2	VGGNet19	80.48	**82.57**	139
3	VGGNet16	**79.67**	77.69	134
4	DenseNet169	73.48	**77.39**	12
5	ResNet101	**72.26**	68.64	23
6	AlexNet	**71.97**	69.98	57
7	ShuffleNet_v2_x2_0	**70.46**	62.99	19
8	InceptionV2	**69.46**	58.44	7
9	GoogLeNet	69	**74.23**	5
10	MobileNetV3	68.71	**69.39**	4
11	ResNet50	**68.28**	63.9	42
12	ShuffleNet_v2_x1_0	63.45	**66.73**	4
13	EfficientNetB0	**61.98**	56.98	4

갑상선 연구

- 실험 4 : 불균형 데이터셋 딜링 스킬(Under-sampling, Over-sampling, Augmentation) 비교 실험

순위	모델	AUC (딜링 O)	AUC (딜링 X)	B_ACC (딜링 O)	B_ACC (딜링 X)
1	DenseNet264	**81.9**	72.86	42.84	**45.64**
2	VGGNet16	77.39	**77.4**	39.08	**42.11**
3	VGGNet19	76.96	**79.47**	35.05	50.3
4	DenseNet169	73.48	**73.55**	**49.72**	38.45
5	AlexNet	**71.97**	58.7	**51.52**	33.33
6	InceptionV2	**69.46**	65.76	**42.75**	39.08
7	GoogLeNet	**67.77**	61.25	**44.19**	33.33
8	MobileNetV3	**63.58**	51.44	33.97	33.24
9	ResNet50	61.19	**66.04**	38.86	**44.24**
10	ShuffleNet_v2_x1_0	**60.39**	57.3	**35.33**	31.07
11	ShuffleNet_v2_x2_0	54.97	**60.37**	32.25	33.97
12	ResNet101	49.61	**64.19**	32.97	**37.36**
13	EfficientNetB0	44.25	**50.7**	31.16	**36.91**

갑상선 연구 : 실험 요약

- DenseNet264, VGGNet19, VGGNet16, DenseNet169 순으로 안정된 성능
- 채널 수 이슈 확인 : 더 많은 정보를 입력해도 아직은 효과적으로 소화하지 못하는 것으로 볼 수 있음 (실험 3)
- 불균형 데이터 딜링 스킬 : 일부 모델에서 효과보았으나, 지속적인 모니터링 필요

갑상선 연구 : 추후 진행사항

- DenseNet264, VGGNet19, VGGNet16, DenseNet169 위주로 실험 진행

  ㄴ(AlexNet, GoogLNet, ResNet까지 가능할듯)

- SOTA 모델 테스트(RGB classification + 의료 이미지 특화)
- 과거 Severity 모델 거의 변형하지 않은 채로 사용