A Feature Selection Study for Medical Image deep learning

Sanghyuck Lee

2021. 06 22.

Outline

- Deep Feature Selection Experiments
- 중견 연구 과제
- 갑상선 연구 과제
- 산학 협력 과제

Performance Measurement

- Performance metric with controlled feature selection:

  Basically, we need Ground Truth

  False Discovery Rate (FDR):

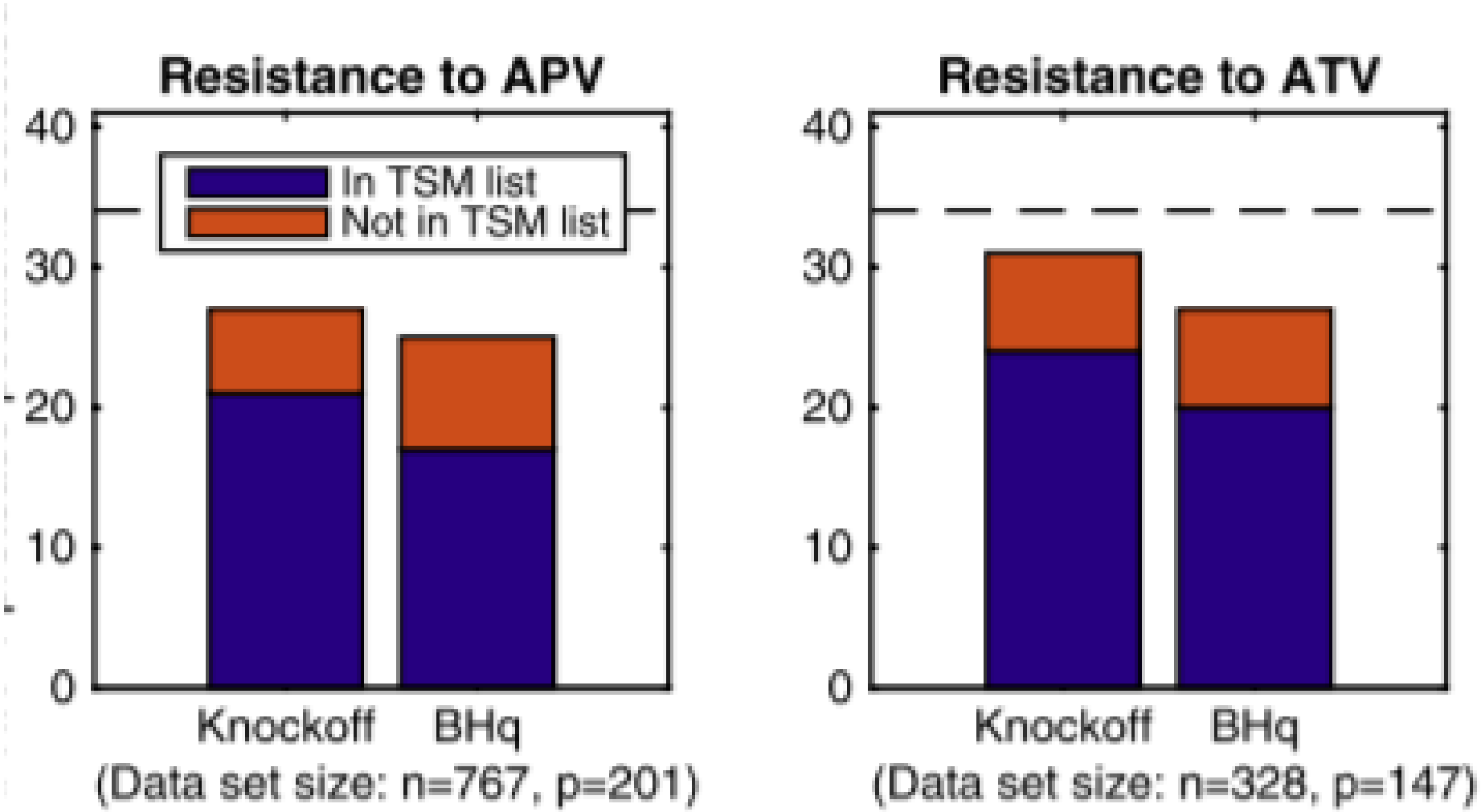
    False among selected features

    FP/(TP+FP)

  Power: Selection ratio of total positives

    TP/(TP+FN)

Performance Measurement



Performance Measurement

p	DeepPINK				MLP			
	Linear		Single-Index		Linear		Single-Index	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power
50	0.046	1	0.13	0.98	0.15	1	0.17	0.89
100	0.047	1	0.08	1	0.048	1	0.056	0.26
200	0.042	0.99	0.042	1	0.11	1	0	0
400	0.022	0.97	0.022	1	0.29	0.95	0	0
600	0.031	0.95	0.046	1	0.17	0.8	0.014	0.013
800	0.048	0.95	0.082	1	0.037	0.62	0.016	0.068
1000	0.023	0.97	0.065	1	0.007	0.4	0.037	0.16
1500	0.007	1	0.065	1	0.002	0.41	0.068	0.25
2000	0.026	0.99	0.098	1	0.023	0.4	0.063	0.35
2500	0.029	0.97	0.067	1	0.21	0.5	0.042	0.35
3000	0.046	0.97	0.051	1	0.11	0.43	0.046	0.31

Traditional VS Controlled feature selection

- Traditional feature selection VS Controlled feature selection

	Traditional feature selection	**Controlled feature selection**
Performance Measure	Accuracy, RMSE, …	FDR, Power
Ground Truth Dataset	Label	Label + Feature

Dataset

- HIV drug resistance dataset (#instances, #features)

1) APV (767, 201)
2) ATV (328, 147)
3) IDV (826, 208)
4) LPV (516, 184)
5) NFV (843, 209)
6) RTV (794, 207)
7) SQV (825, 208)
8) 3TC (632, 292)
9) ABC (627, 294)
10) AZT (629, 292)
11) D4T (629, 293)
12) DDI (631, 292)
13) TDF (352, 218)
14) DLV (731, 311)
15) EFV (733, 318)
16) NVP (745, 319)

Performance

	Proposed		DeepPink
Dataset	**Power**	**FDR**	**FDR**
APV	0.382	0.417	**0**
ATV	0.191	0.308	**0.278**
IDV	0.625	0.646	**0.24**
LPV	0.382	0.4	**0.217**
NFV	0.607	0.507	**0.263**
RTV	0.714	0.524	**0.182**
SQV	0.679	0.525	**0.133**
3TC	0	1	**0.2**

	Proposed		DeepPink
Dataset	**Power**	**FDR**	**FDR**
ABC	0.143	0.333	**0.25**
AZT	0.452	0.683	**0.278**
D4T	0.452	0.296	**0.167**
DDI	0.476	0.5	**0.2**
TDF	0	1	**0.167**
DLV	0.31	0.75	.
EFV	0.818	0.842	.
NVP	0.909	0.828	.

중견 연구 제안서 작성

- 가제:

  한글: 의료 분야내 주요 관심 구조물 세그멘테이션을 위한 사전훈련 최신 딥러닝 모델 개발

  영문: Development of Pre-trained Deep Learning Model for Segmentation of Structures of Major Interest in the Medical Field

- 연구 목표:

  본 과제에서는 몇몇 주요한 해부학적 구조에 대해서 세계 최고 수준의 성능을 갖는 사전학습된 세그멘테이션 모델을 배포함으로써, 다양한 의료분야 응용에 성공적으로 기여하는 것을 최종 목표로 하며, 이를 위해 다음과 같은 연차별 연구 목표를 설정하여 추진한다.

갑상선 연구

- 완료 현황: 각각 1237 환자

유형	눈 (L)	눈 (R)	시신경 (L)	시신경 (R)	내직근 (L)	내직근 (R)	외직근 (L)	외직근 (R)	상직근 (L)	상직근 (R)	하직근 (L)	하직근 (R)	상안검 (L)	상안검 (R)	지방 (L)	지방 (R)	눈물주머니 (L)	눈물주머니 (R)	안와 (L)	안와 (R)	평균
AX 1	0.91	0.87	0.72	0.79	0.91	0.94	0.93	0.94							0.42	0.45					0.79
AX 2	0.91	0.89	0.75	0.72	0.93	0.93	0.93	0.9							0.56	0.64					0.82
AX 3	0.96	0.96	0.95	0.94	0.95	0.96	0.94	0.95							0.86	0.89					0.94
AX 4	0.94	0.94	0.82	0.84	0.82	0.84	0.79	0.75							0.63	0.56					0.79
AX 5	0.95	0.95	0.98	0.98	0.9	0.93	0.96	0.95							0.45	0.44					0.85
AX 6																	0.84	0.89			0.87
CO 1	0.96	0.96																			0.96
CO 2			0.89	0.89	0.91	0.91	0.91	0.89	0.86	0.9	0.91	0.88									0.90
CO 3																			0.63	0.83	0.73
SA 1		1		0.99						1		0.99		1							1.00
SA 2	1		0.99						1		0.99		1								1.00
계																					0.87

0% 100%

산학 협력 프로젝트

- 제출 실적
  - SW 제작 (산학 1), 국제학술대회 논문 1편 이상 게재 (산학 2), 국내 특허출원 1건 이상 (산학 2)
- 현재 진행 사항
  - SW 등록: 진행중
  - 특허 출원: 기존 논문 번역본 제작 중
  - 투고 저널 리스트업 완료

References

Barber, Rina Foygel, and Emmanuel J. Candès. "Controlling the false discovery rate via knockoffs." *Annals of Statistics* 43.5 (2015): 2055-2085.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995): 289-300.

Chen, Jiajie, Anthony Hou, and Thomas Y. Hou. "Some analysis of the knockoff filter and its variants." *arXiv preprint arXiv:1706.03400* (2017).

Jordon, James, Jinsung Yoon, and Mihaela van der Schaar. "KnockoffGAN: Generating knockoffs for feature selection using generative adversarial networks." *International Conference on Learning Representations*. 2018.

Sarkar, Sanat K., and Cheng Yong Tang. "Adjusting the Benjamini-Hochberg method for controlling the false discovery rate in knockoff assisted variable selection." *arXiv preprint arXiv:2102.09080* (2021).