Histopathology case report machine learning for disease prediction and search on social media and PubMed

Andrew J. Schaumberg[1,2,3,*] Wendy Juarez[3,4,ɣ], Sarah J. Choudhury[3,4,ɣ], Laura G. Pastrián[5,β], Bobbi S. Pritt[6,β], Mario Prieto Pozuelo[7,β], Ricardo Sotillo Sánchez[8,β], Khanh Ho[9,β], Nusrat Zahra[10,β], Betul Duygu Sener[11,β], Stephen Yip[12,β], Bin Xu[3,β], Srinivas Rao Annavarapu[14,β], Aurélien Morini[15,β], Karra A. Jones[16,β], Kathia Rosado-Orozco[17,β], Sanjay Mukhopadhyay[18,β], Carlos Miguel[19,β], Hongyu Yang[20,β], Yale Rosen[21,β], Rola H. Ali[22,β], Olaleke Folaranmi[23,β], Jerad M. Gardner[24,β], Corina Rusu[25,β], Celina Stayerman[26,β], John Gross[27,β], Dauda Suleiman[28,β], S. Joseph Sirintrapun[29], Mariam Aly[30,31,δ,*], Thomas J. Fuchs[2,29,δ,*]

Weill Cornell Medicine
Graduate School
A partnership with the Sloan Kettering Institute
High School Science Immersion Program

[1] Memorial Sloan Kettering Cancer Center and the Tri-Institutional Training Program in Computational Biology and Medicine, NY, USA, [2] Weill Cornell Graduate School of Medical Sciences, NY, USA, [3] Weill Cornell High School Science Immersion Program, [4] Manhattan/Hunter Science High School, NY, USA, [5] Hospital Universitario La Paz, Madrid, Spain, [6] Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN, USA, [7] Hospital Universitario HM Sanchinarro, Laboratorio de Dianas Terapéuticas, Madrid, Spain, [8] Virgen de Altagracia Hospital, Manzanares, Spain, [9] Centre Hospitalier de Mouscron, Belgium, [10] Allama Iqbal Medical College, Lahore, Pakistan, [11] Konya Training and Research Hospital, Konya, Turkey, [12] BC Cancer, British Columbia, Canada, [13] Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada, [14] Royal Victoria Infirmary, Department of Cellular Pathology, Newcastle upon Tyne, England, UK, [15] Université Paris Est Créteil, Faculté de médecine de Créteil, France, [16] University of Iowa, Department of Pathology, IA, USA, [17] HRP Labs, San Juan, Puerto Rico, USA, [18] Cleveland Clinic, Department of Pathology, Cleveland, OH, USA, [19] Centro Médico de Asturias, Department of Pathology, Oviedo, Spain, [20] St Vincent Evansville Hospital, Department of Pathology, Evansville, IN, USA, [21] SUNY Downstate Medical Center, Department of Pathology, NY, USA, [22] Kuwait University, Department of Pathology, Faculty of Medicine, Kuwait, [23] University of Ilorin Teaching Hospital, Department of Pathology, Nigeria, [24] University of Arkansas for Medical Sciences, Department of Pathology, Little Rock, AK, USA, [25] Augusta Hospital, Department of Pathology, Bochum, Germany, [26] Laboratorio TechniPath, Ceutec, Honduras, [27] Mayo Clinic, Bone and Soft Tissue and Surgical Pathology, Rochester, MN, USA, [28] Abubakar Tafawa University Teaching Hospital, Department of Histopathology, Bauchi, Nigeria, [29] Memorial Sloan Kettering Cancer Center, Department of Pathology, NY, USA, [30] Columbia University, Department of Psychology, NY, USA, [31] Affiliate Member of the Zuckerman Mind Brain Behavior Institute, Columbia University

ɣEqual contribution. β Generously donated pathology cases. δ Principal Investigators of this work. *Contact ajs625@cornell.edu ma3631@columbia.edu fuchst@mskcc.org

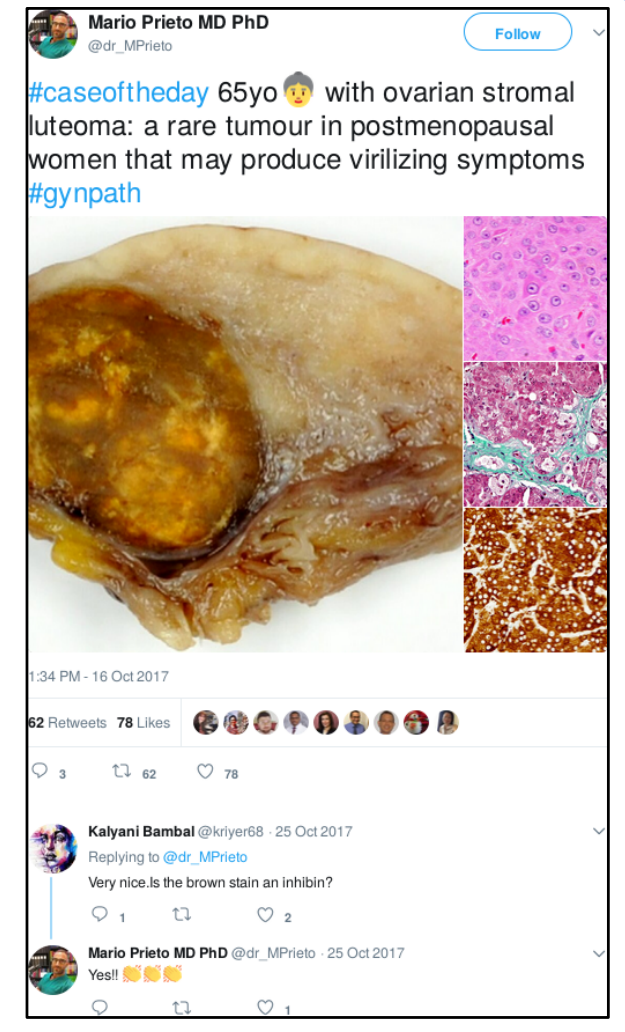
bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

Problem

Large-scale annotated image datasets, e.g. ImageNet, are essential in testing new methods for natural vision tasks.

Diagnosing cancer is a challenging vision task requiring highly-trained pathologists.

Pathologists seek additional opinions for challenging cases, but on-site colleagues are not always immediately available.



Mario Prieto MD PhD @dr_MPrieto Follow

#caseoftheday 65yo with ovarian stromal luteoma: a rare tumour in postmenopausal women that may produce virilizing symptoms #gynpath

1:34 PM - 16 Oct 2017

62 Retweets 78 Likes

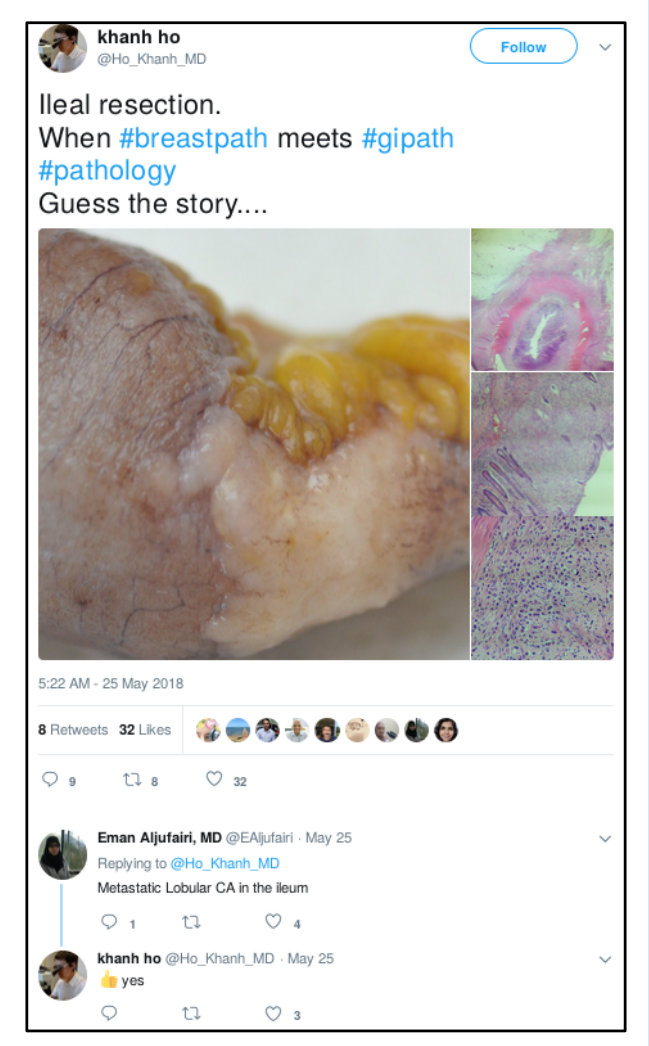
Kalyani Bambal @kriyer68 · 25 Oct 2017
Replying to @dr_MPrieto
Very nice.Is the brown stain an inhibin?

Mario Prieto MD PhD @dr_MPrieto · 25 Oct 2017
Yes!!

Solution

We obtained a large annotated pathology dataset from social media and PubMed, and implemented a bot (@pathobot) to search and classify.

Mentioning this bot alongside a pathology image leads it to search for similar cases and predict disease states via machine learning.



khanh ho @Ho_Khanh_MD Follow

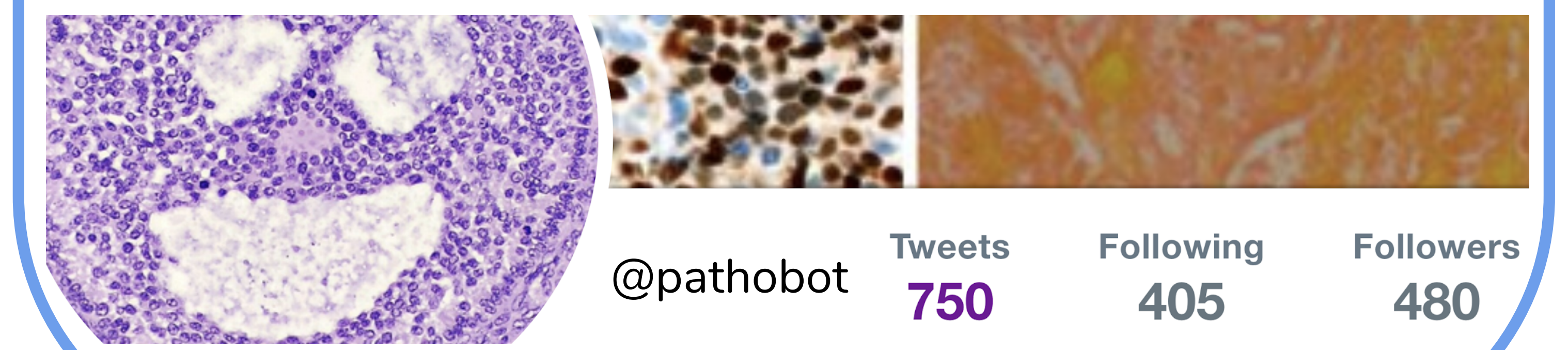
Ileal resection.
When #breastpath meets #gipath #pathology
Guess the story....

5:22 AM - 25 May 2018

8 Retweets 32 Likes

Eman Aljufairi, MD @EAljufairi · May 25
Replying to @Ho_Khanh_MD
Metastatic Lobular CA in the ileum

khanh ho @Ho_Khanh_MD · May 25
yes

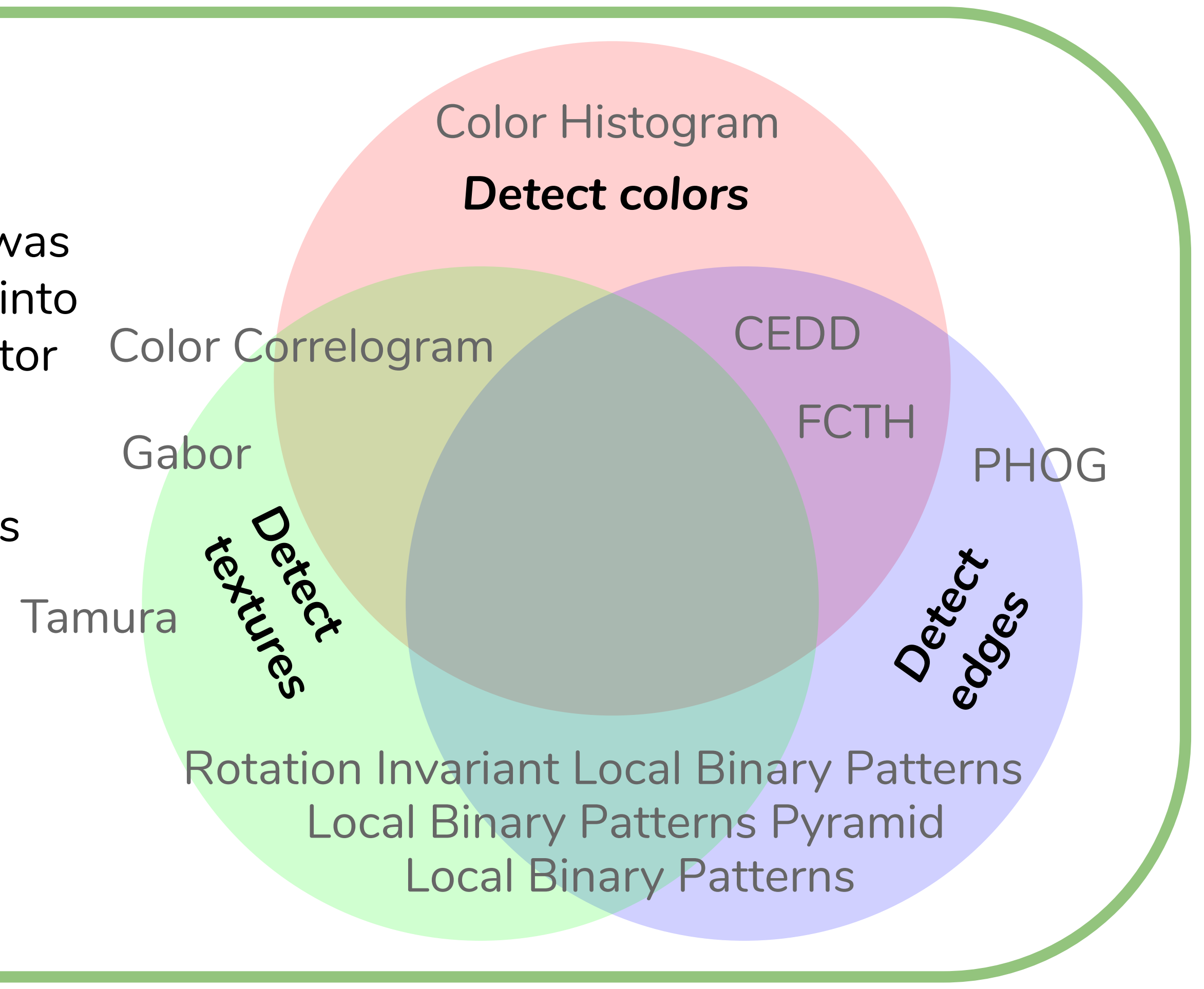


@pathobot

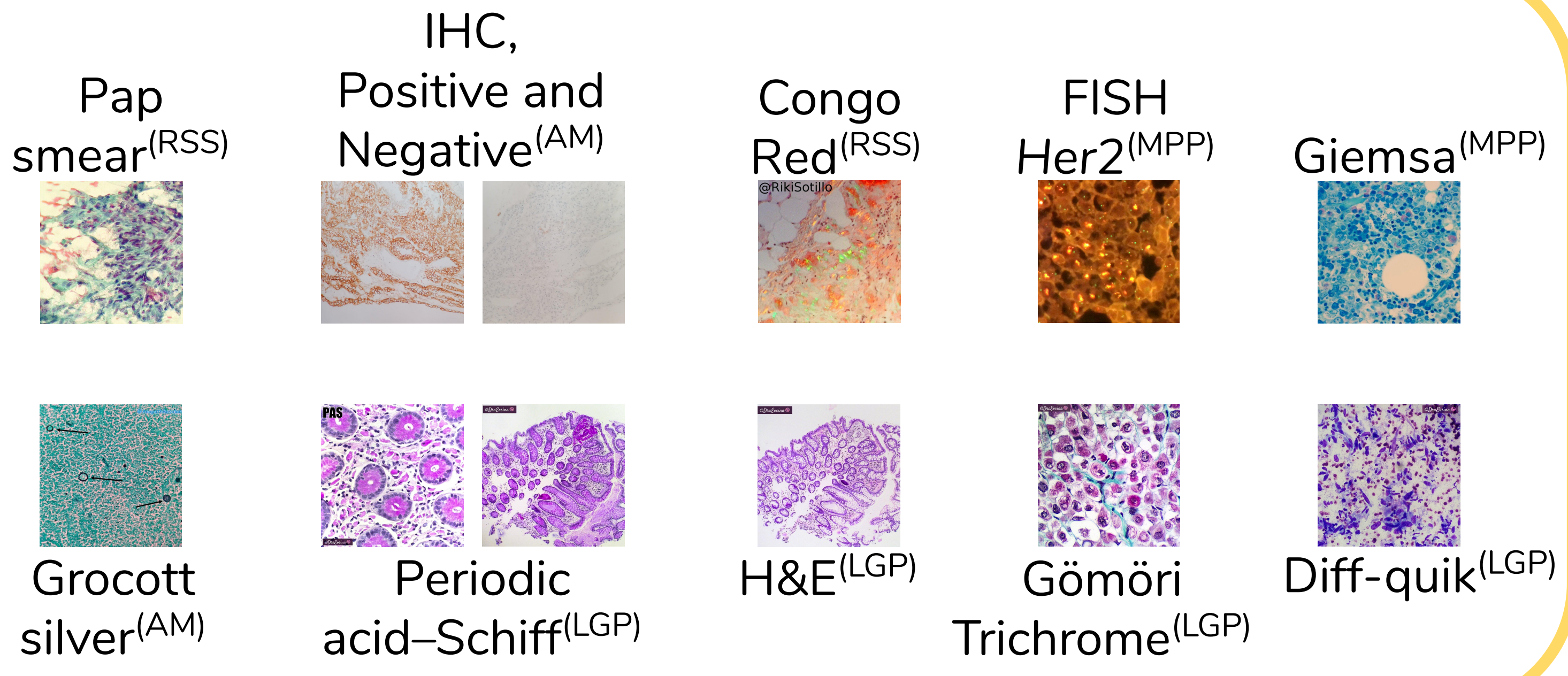
Tweets	Following	Followers
750	405	480

Features

Each image was transformed into a feature vector of 2412 numbers. These vectors are used for machine learning, prediction, and search.

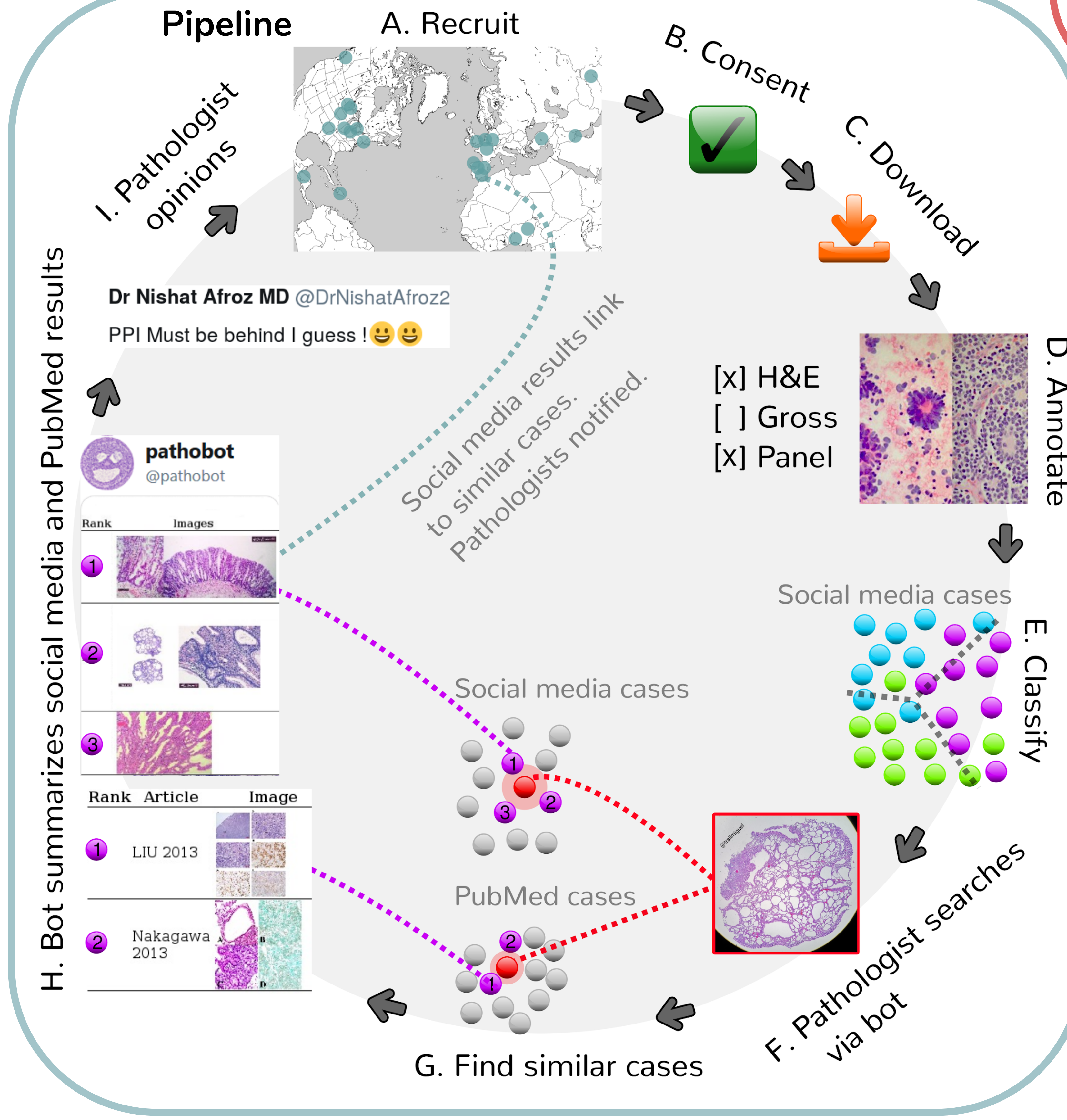


Technique diversity



Pap smear(RSS)
IHC, Positive and Negative(AM)
Congo Red(RSS)
FISH Her2(MPP)
Giemsa(MPP)
Grocott silver(AM)
Periodic acid–Schiff(LGP)
H&E(LGP)
Gömöri Trichrome(LGP)
Diff-quik(LGP)

Pipeline



A. Recruit
B. Consent
C. Download
D. Annotate
[x] H&E
[ ] Gross
[x] Panel
E. Classify
Social media cases
F. Pathologist searches via bot
G. Find similar cases
Social media cases
PubMed cases
H. Bot summarizes social media and PubMed results
I. Pathologist opinions
Social media results link to similar cases. Pathologists notified.

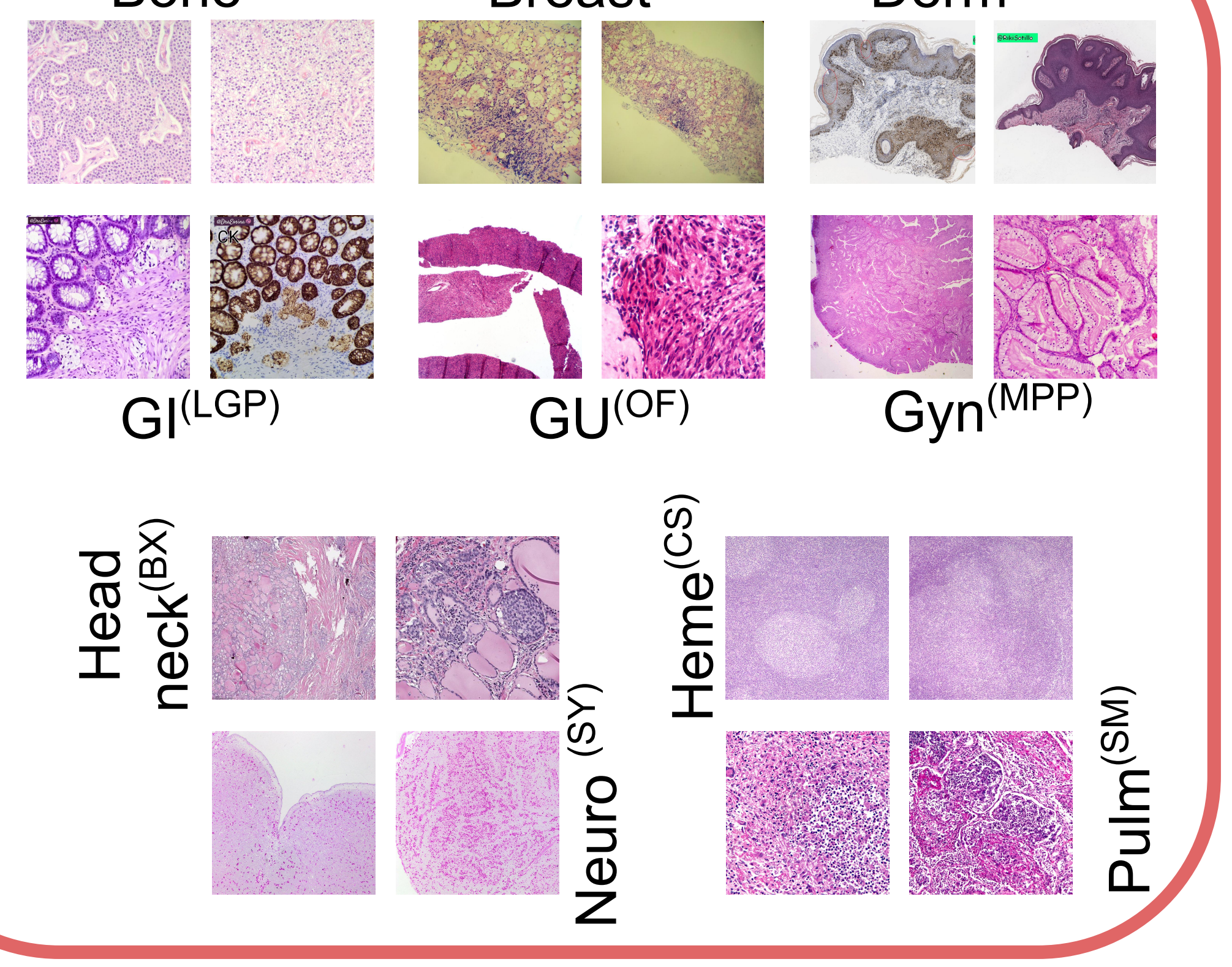
Dr Nishat Afroz MD @DrNishatAfroz2
PPI Must be behind I guess !😃😃

pathobot @pathobot

Rank	Images
1	
2	
3	

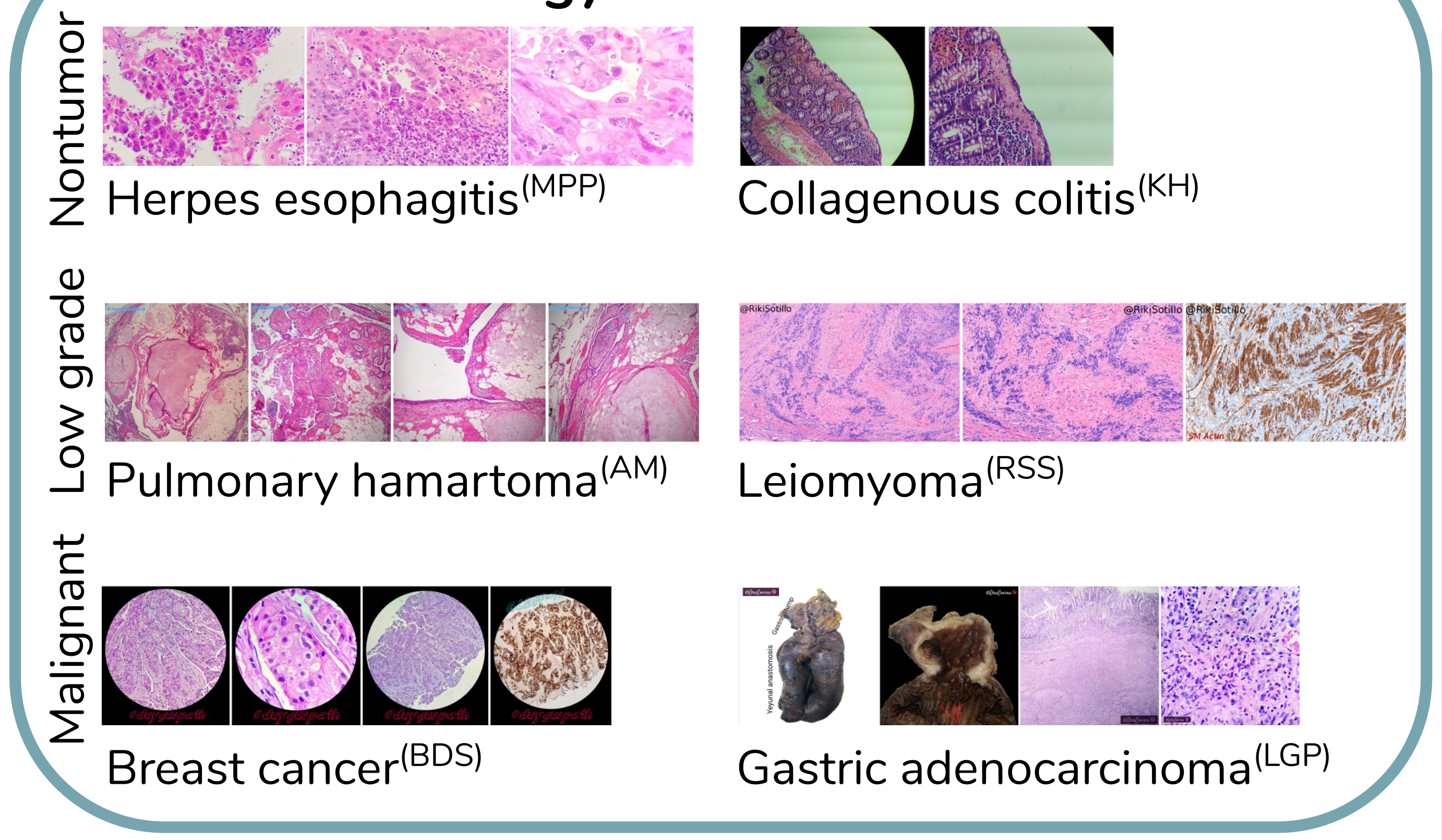
Rank	Article	Image
1	LIU 2013	
2	Nakagawa 2013	

Histopathology types



Bone(CS)
Breast(KH)
Derm(RSS)
GI(LGP)
GU(OF)
Gyn(MPP)
Head neck(BX)
Neuro(SY)
Heme(CS)
Pulm(SM)

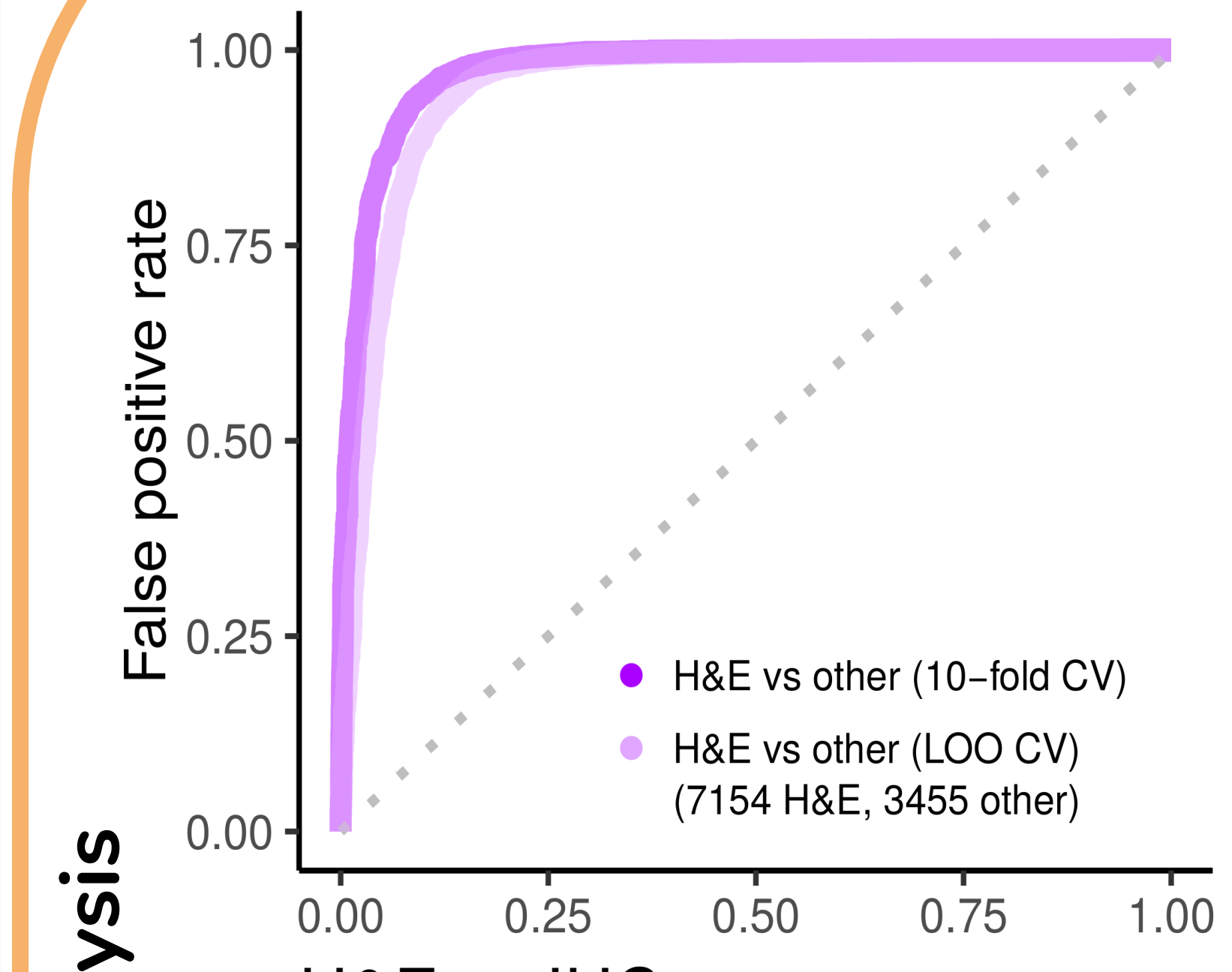
Pathology disease states



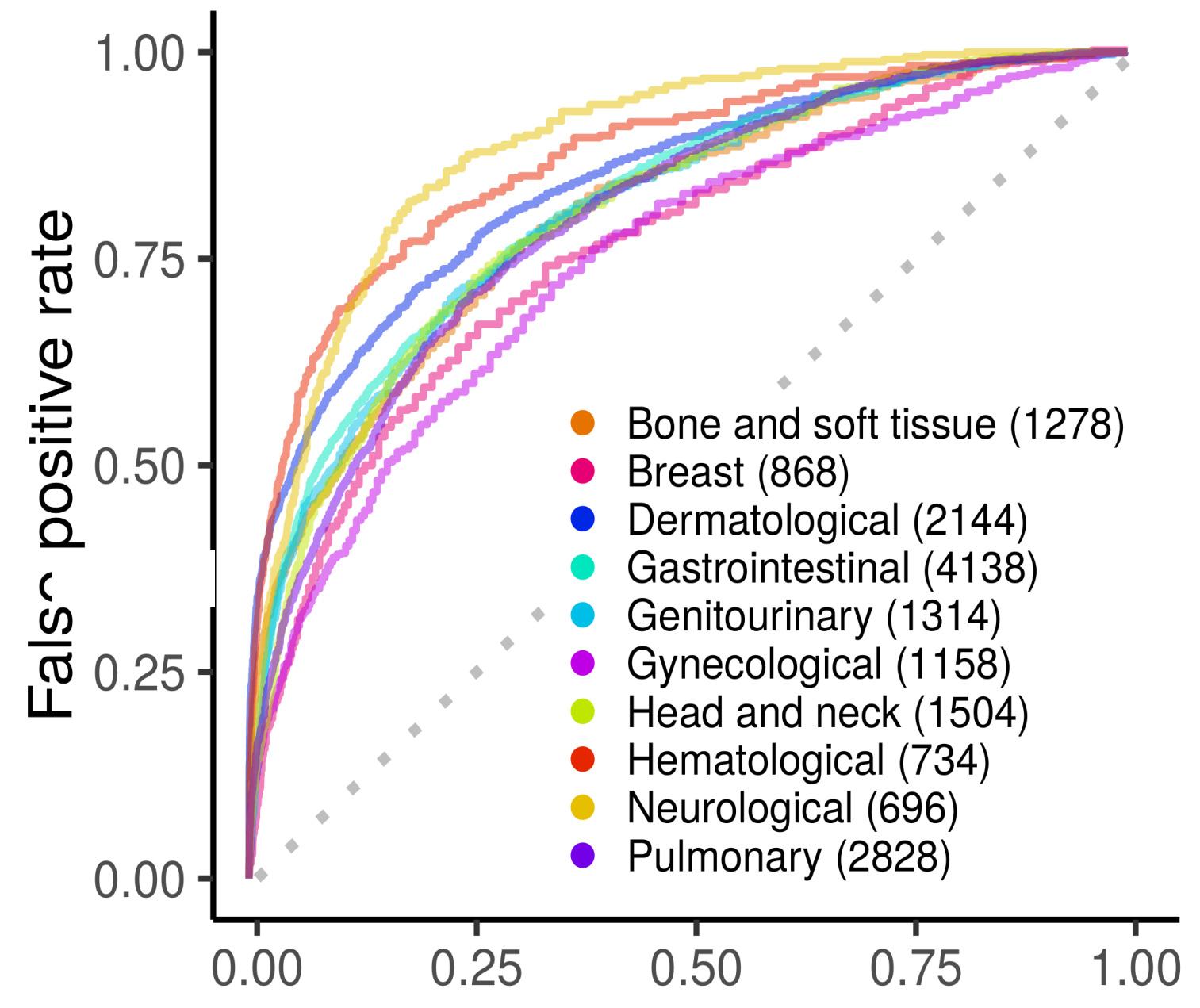
Nontumor
Herpes esophagitis(MPP)
Collagenous colitis(KH)
Low grade
Pulmonary hamartoma(AM)
Leiomyoma(RSS)
Malignant
Breast cancer(BDS)
Gastric adenocarcinoma(LGP)

Analysis

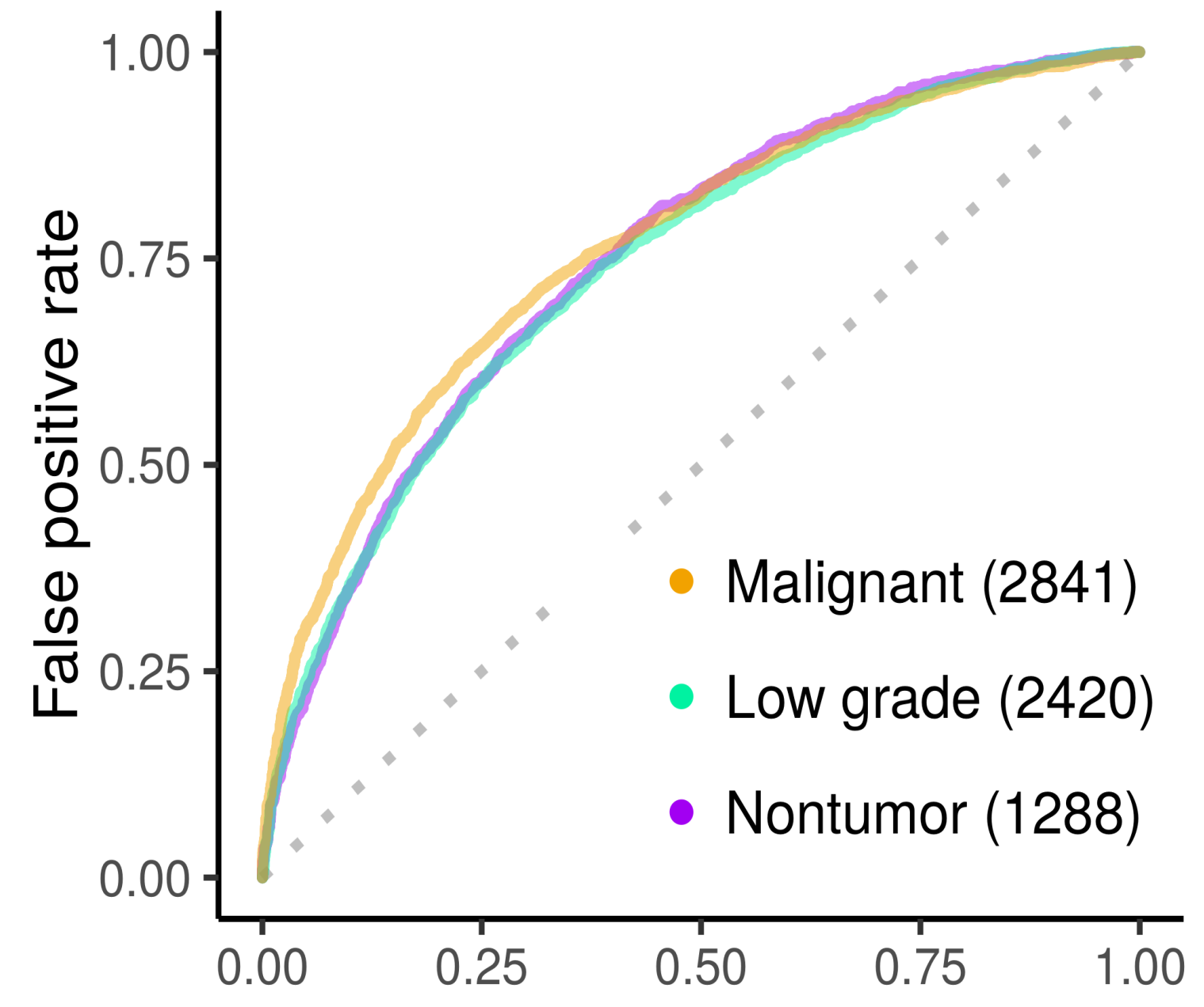
H&E vs others
10−fold, Leave−one−pathologist−out



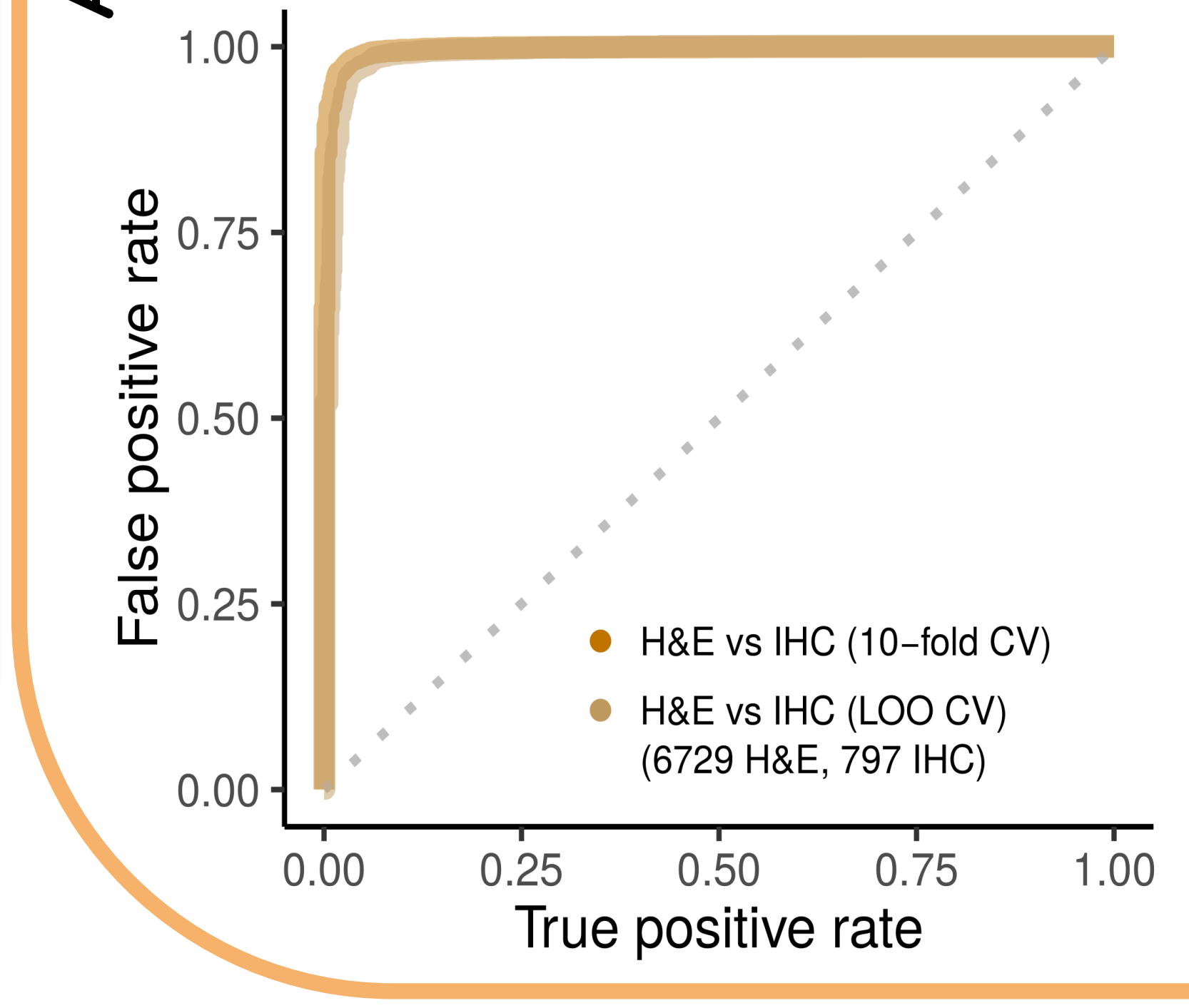
Tissue type
10−fold CV



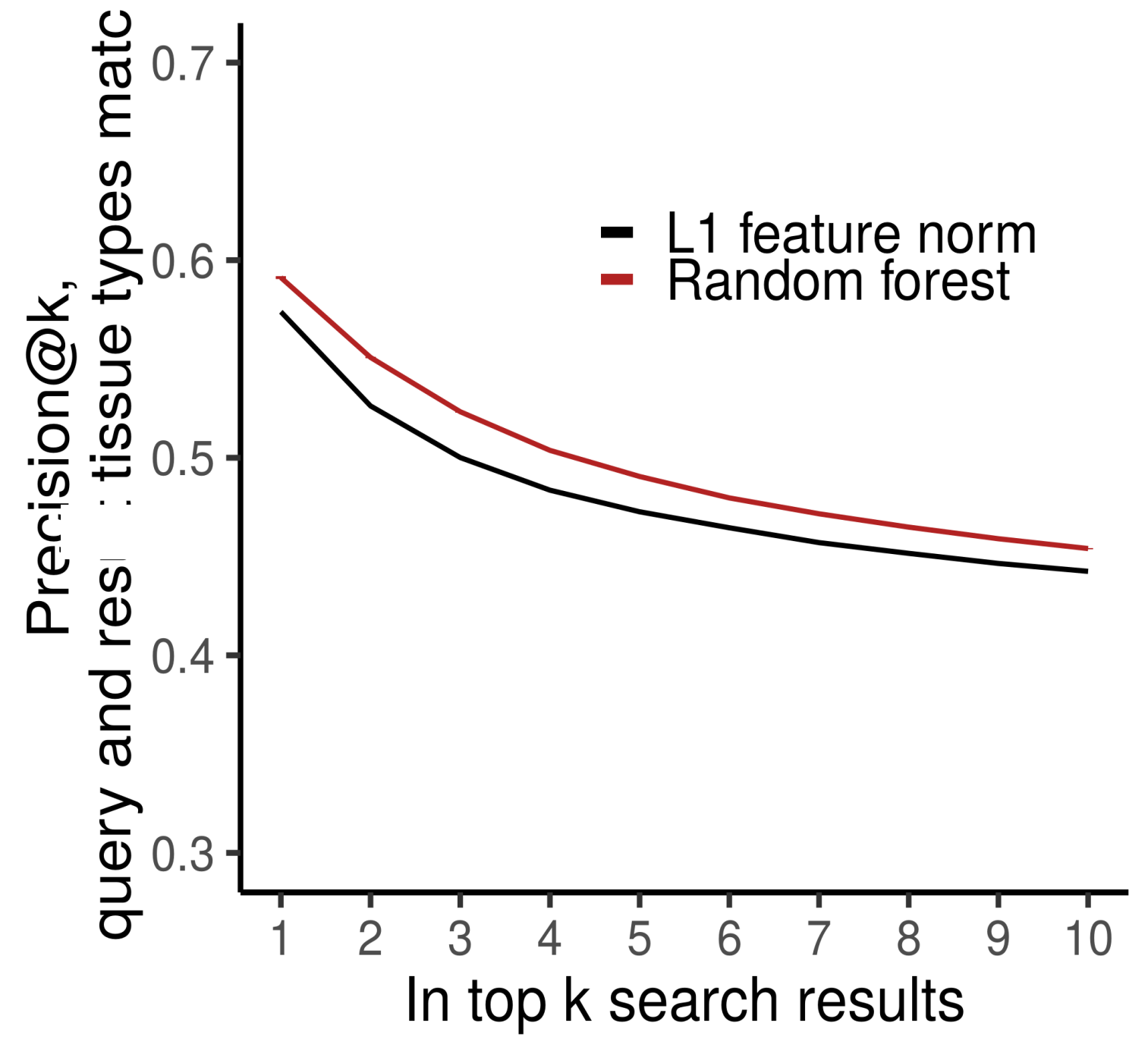
Disease state
10−fold CV



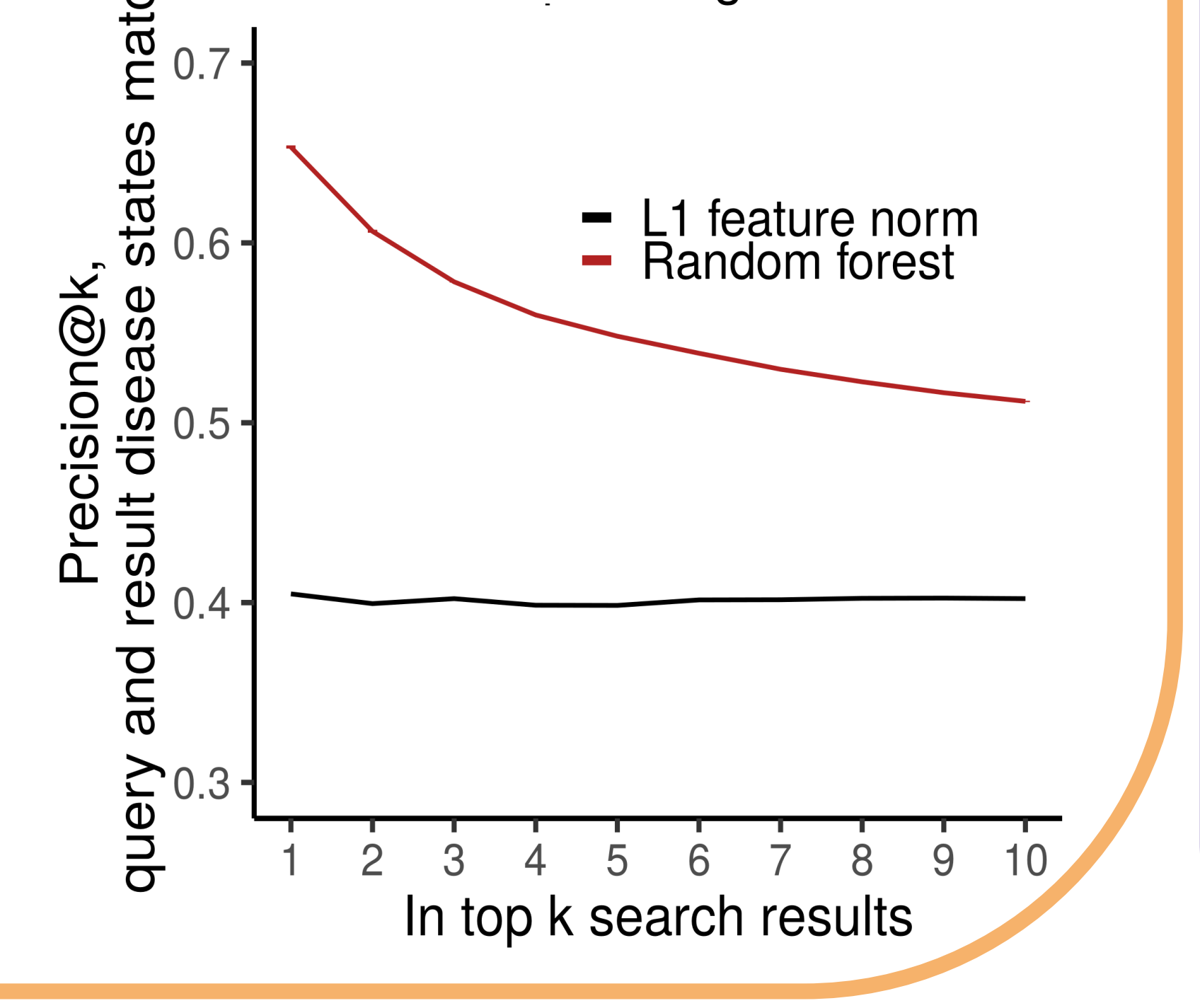
H&E vs IHC
10−fold, Leave−one−pathologist−out



Tissue type precision@k
Leave−one−pathologist−out



Disease state precision@k
Leave−one−pathologist−out



Conclusions

We constructed a dataset of 6,351 Twitter messages from 25 pathologists worldwide, and used it for machine learning. This was supplemented with 113,161 H&E stained images from PubMed, for sampling rare diseases.

Machine learning on these images, in conjunction with pathologist dialogue, facilitates diagnoses and decisions about next steps in patient care. This is accomplished by connecting pathologists all over the world, searching for similar cases, and generating predictions about disease states in shared images.

To our knowledge, this is the first use of social media data for pathology case search, and the first pathology study prospectively tested in full public view on social media. We expect our project to cultivate a more connected world of physicians and improve patient care worldwide.