Deep neural network models in computational histopathology

A second computed-diagnosis opinion for pathologists

PhD candidate: Caredio Davide

Technical Journal Club: 13.04.2021

Aguzzi's Lab

INTRODUCTION

Histopathological images contain rich phenotypic information that can be used to monitor underlying mechanisms contributing to disease progression and patient survival outcomes.

Pathological analysis under the microscope is inherently subjective in nature. Differences in visual perception and clinical training can lead to inconsistencies in diagnostic and prognostic opinions.

Even experienced pathologists are prone to misdetect features and make errors due to the enormous amount of tissue to analyse

The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy

Fadi Brimo [1], Luciana Schultz, Jonathan I Epstein

Affiliations + expand

PMID: 20478583 DOI: 10.1016/j.juro.2010.03.021

March 17, 2015

Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens

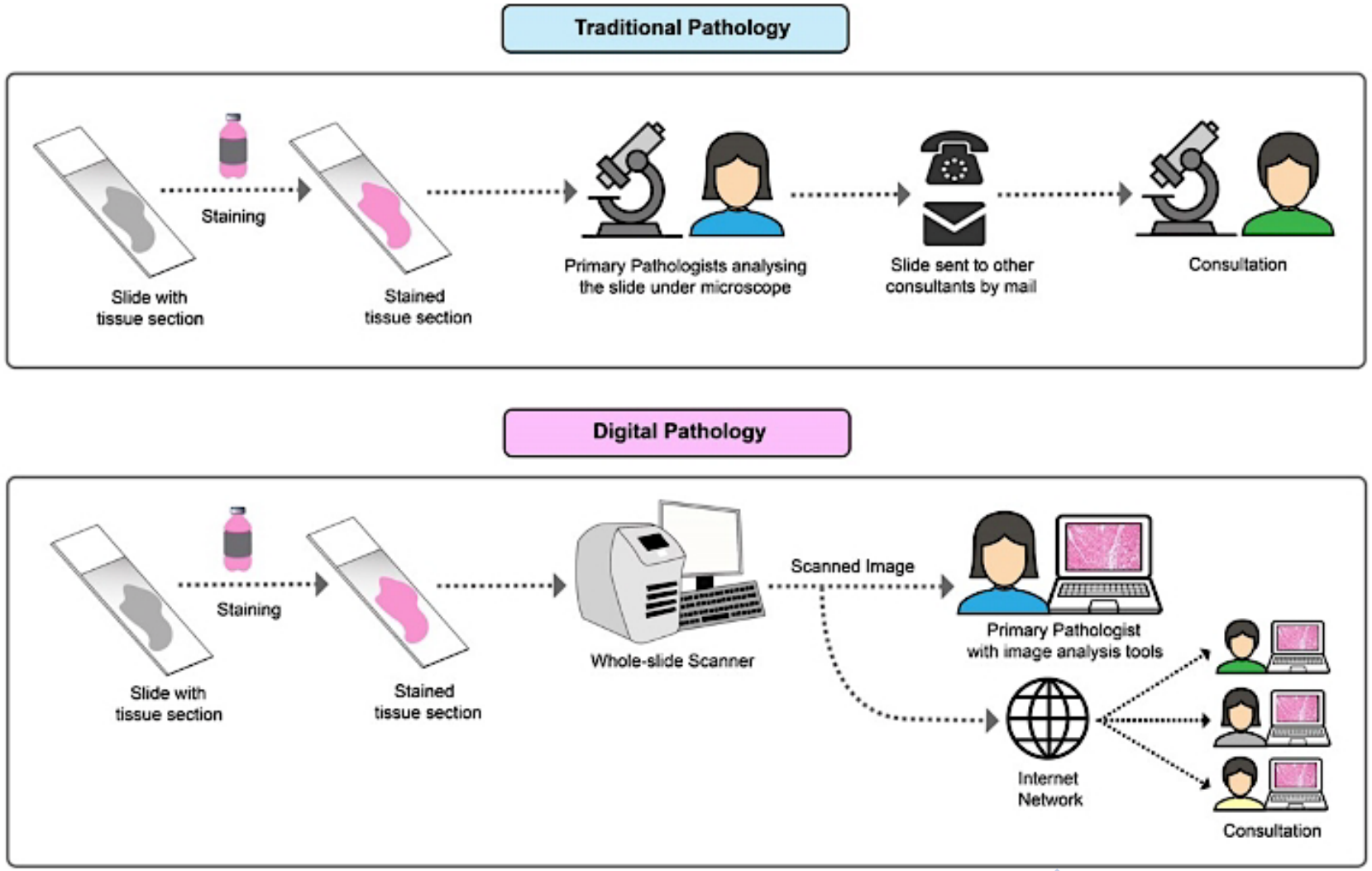
Joann G. Elmore, MD, MPH[1]; Gary M. Longton, MS[2]; Patricia A. Carney, PhD[3]; et al

» Author Affiliations | Article Information

JAMA. 2015;313(11):1122-1132. doi:10.1001/jama.2015.1405

Qualified diagnosis requires peer review and consensus.

In the 1990s, hospitals adopted sub-micron-level resolution tissue scanners that capture gigapixel **whole-slide images** (WSI).



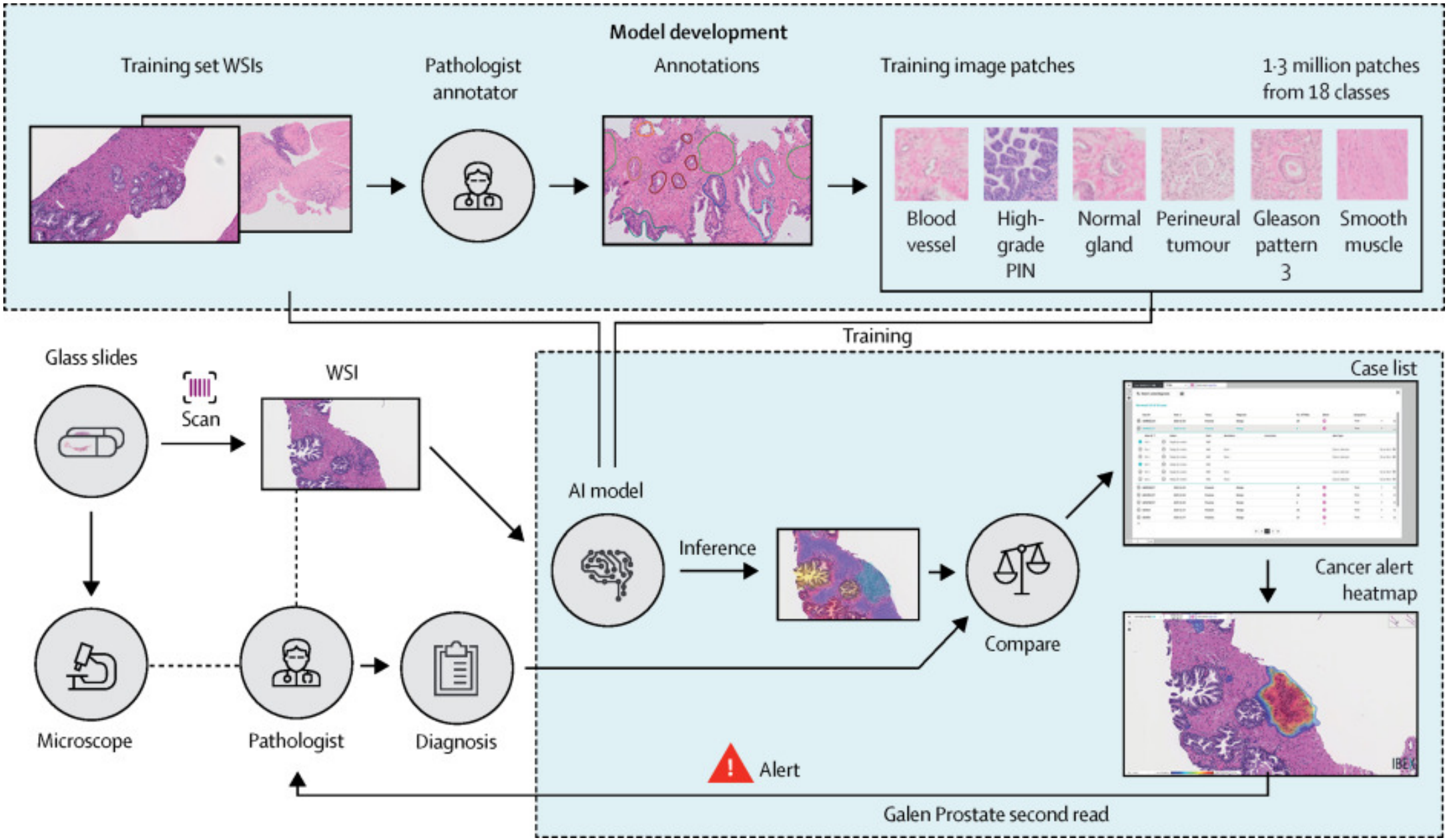
Few areas (radiology, pathology, ophthalmology, and dermatology) have received substantial attention owing to such availability of highly structured images.

Physicians assign pathological scores to digital slides depending on the disease phenotypes which can be detected by AI software

Here, **Deep Learning** (DL) can support critical medical tasks, including diagnostics, prognostication of outcomes and treatment response, pathology segmentation, disease monitoring, and so forth.

DL has been introduced into diagnostic workflows with the aim of **secondary diagnostic opinion**.

Overview of the algorithm and clinical deployment of the Galen Prostate second read system



Pantanowitz et al., 2020. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images. The Lancet.

AI applications

- High content screening
- Cytology and **histopathology**
- Time-lapse image analysis

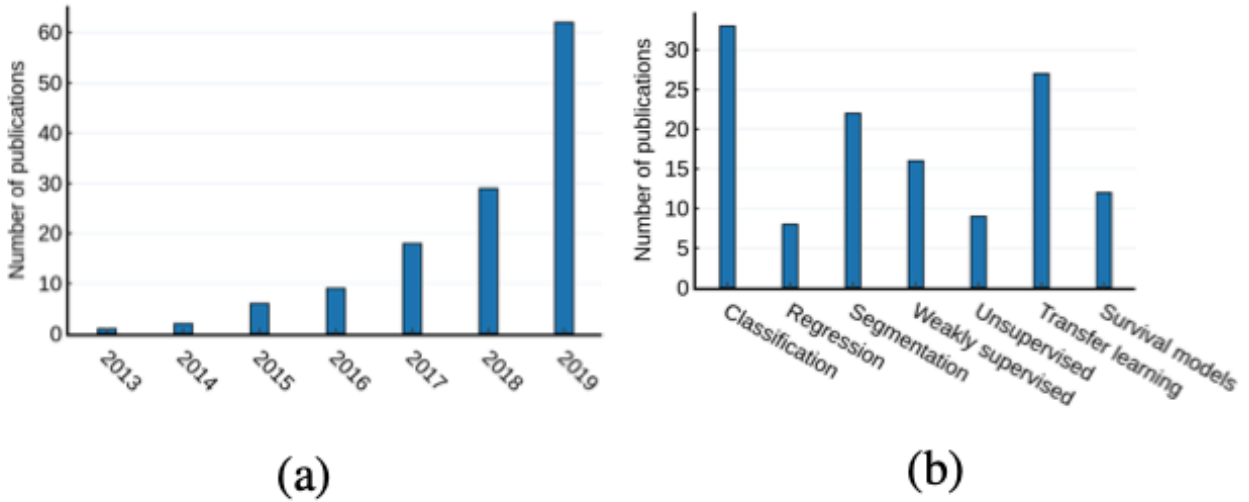
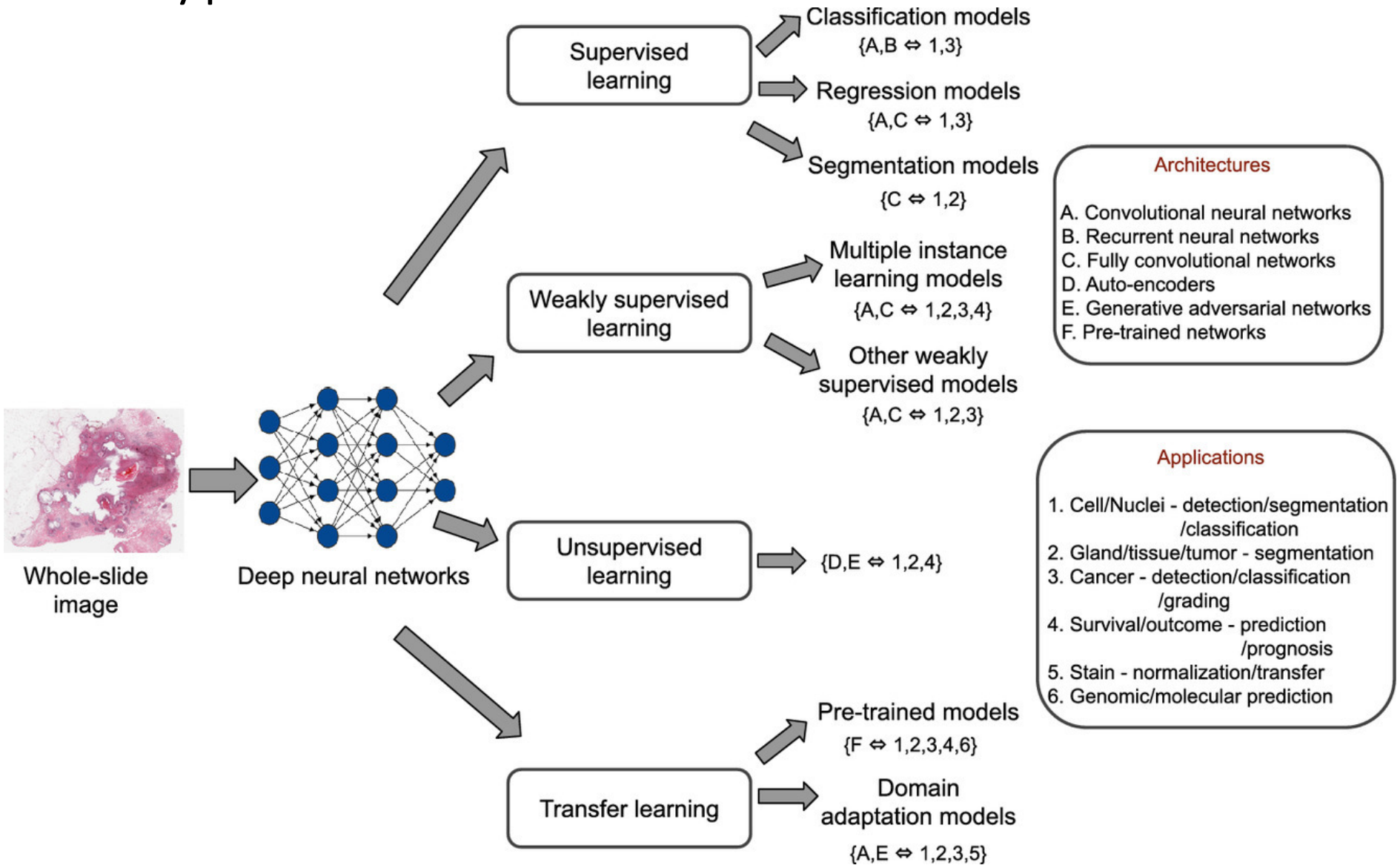


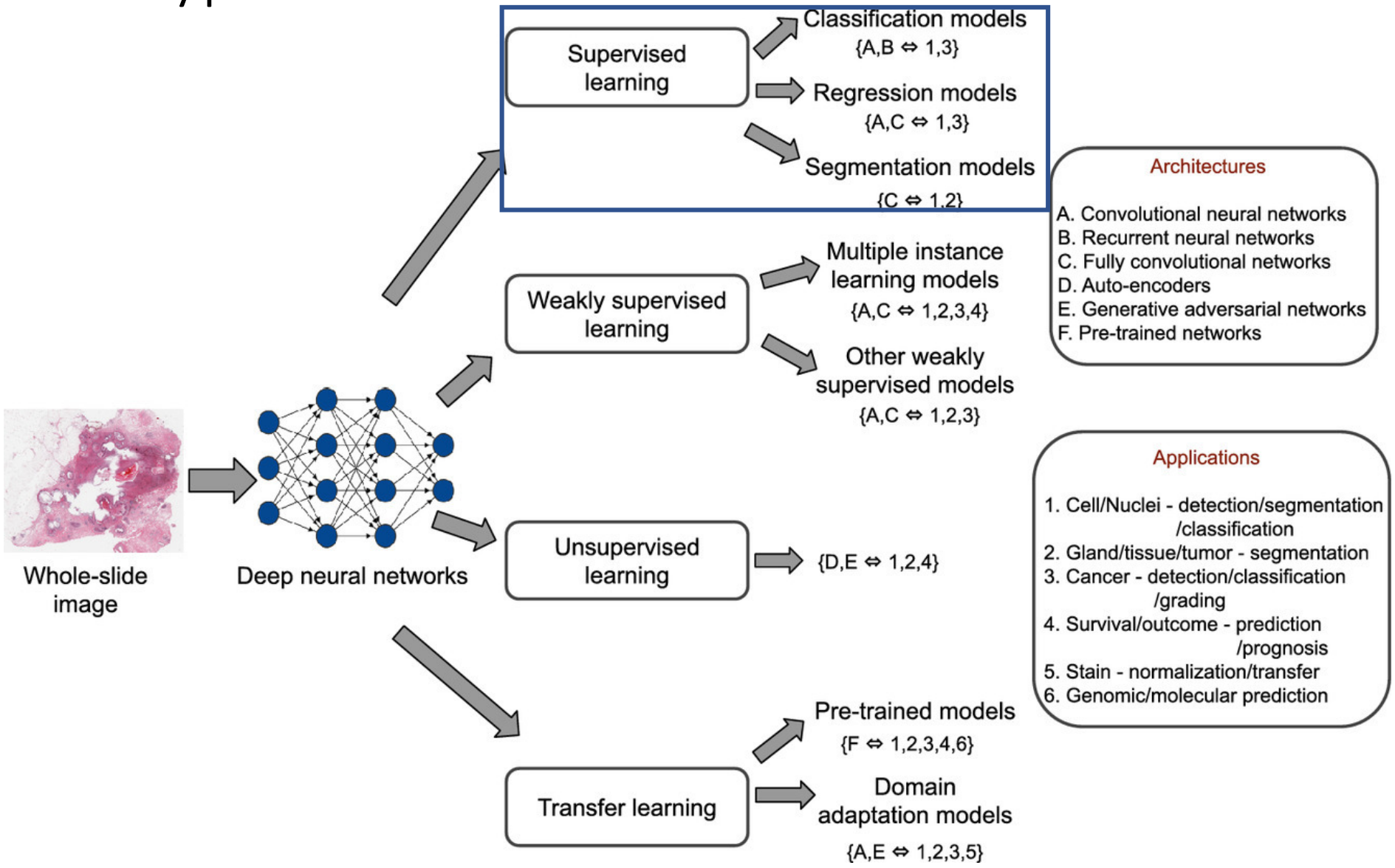
Fig. 1: (a) An overview of numbers of papers published from January 2013 to December 2019 in deep learning based computation histopathology surveyed in this paper. (b) A categorical breakdown of the number of papers published in each learning schemas.

Chetan L. Srinidhi et al., *Deep neural network models for computational histopathology: A survey*, Medical Image Analysis (2020)

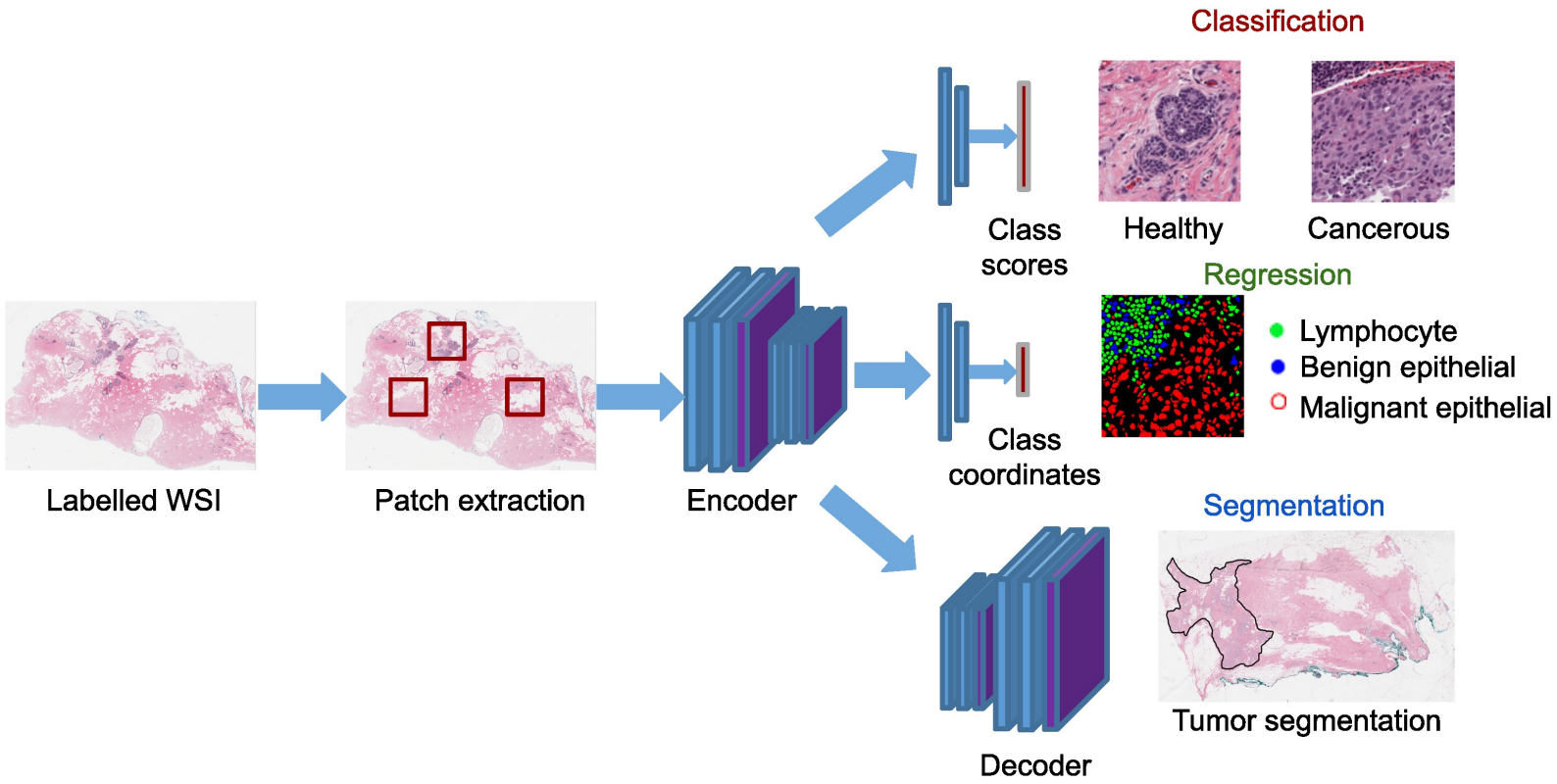
Different type of AI software



Different type of CNN



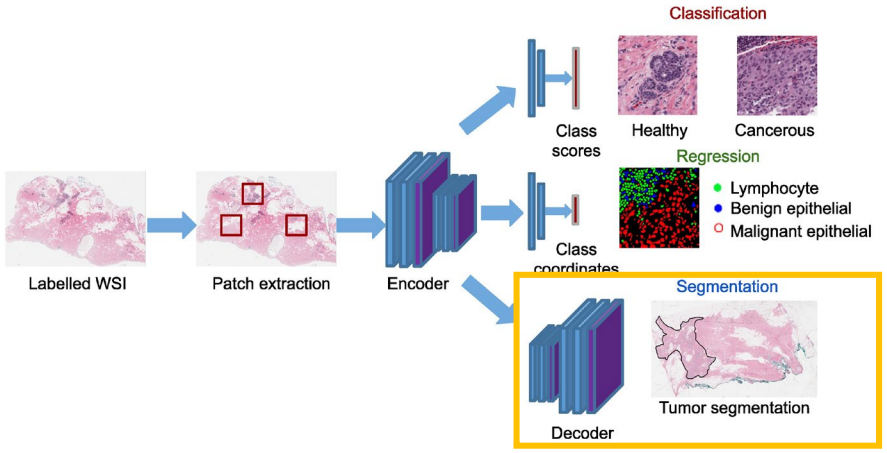
Supervised Learning



- object detection or image classification tasks
- predicting the position of objects (e.g., cells or nuclei)
- predicting cancer severity score by enforcing topological/spatial constraints
- used to solve pathological volumetric changes in medical imaging scenarios.

Supervised Learning

- There is the need to **annotate** a huge datasets before using the network. It is time-consuming for pathologists.
- Outsourcing labelling to non-experts can, however, lead to subjective and inconsistent labels and conventional DL models may find it challenging to train with noisy annotations.
- An alternative approach is to make use of expert advice by providing feedback for annotating rare and challenging cases.
- A possibility is to use a set of precomputed annotations from an automated system, and correct only those labels with inconsistent markings under expert supervision (Marzahl et al., 2019)

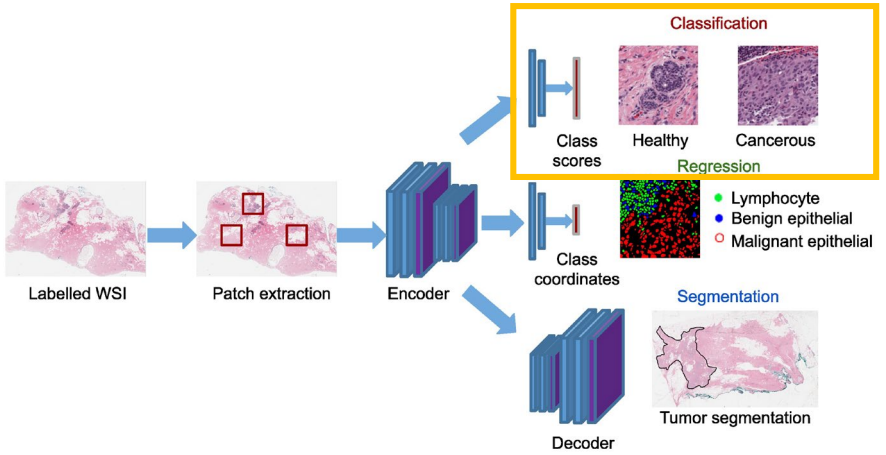


nature COMMUNICATIONS

https://doi.org/10.1038/s41467-019-10212-1 OPEN

Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline

Ziqi Tang[1,2], Kangway V. Chuang[1], Charles DeCarli[3], Lee-Way Jin[4], Laurel Beckett[5], Michael J. Keiser[1] & Brittany N. Dugger[6]



ARTICLES
https://doi.org/10.1038/s42256-019-0052-1
nature machine intelligence

Corrected: Publisher Correction; Publisher Correction

Pathologist-level interpretable whole-slide cancer diagnosis with deep learning

Zizhao Zhang[1], Pingjun Chen[2], Mason McGough[2], Fuyong Xing[3], Chunbao Wang[4], Marilyn Bui[5], Yuanpu Xie[2], Manish Sapkota[6], Lei Cui[2], Jasreman Dhillon[5], Nazeel Ahmad[7], Farah K. Khalil[5], Shohreh I. Dickinson[5], Xiaoshuang Shi[2], Fujun Liu[6], Hai Su[2], Jinzheng Cai[2] and Lin Yang[2*]

Supervised Learning – Segmentation

Segmentation of histological tissue components is an essential pre-requisite for obtaining reliable morphological measurements.

- Pixel-level delineation of object contour or the whole interior of the object of interest is required.
- Best model so far achieved is the U-Net DCNN for histopathology

nature > nature communications > articles > article

Article | Open Access | Published: 15 May 2019

Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline

Ziqi Tang, Kangway V. Chuang, Charles DeCarli, Lee-Way Jin, Laurel Beckett, Michael J. Keiser & Brittany N. Dugger

Nature Communications **10**, Article number: 2173 (2019) | Cite this article

13k Accesses | **22** Citations | **168** Altmetric | Metrics

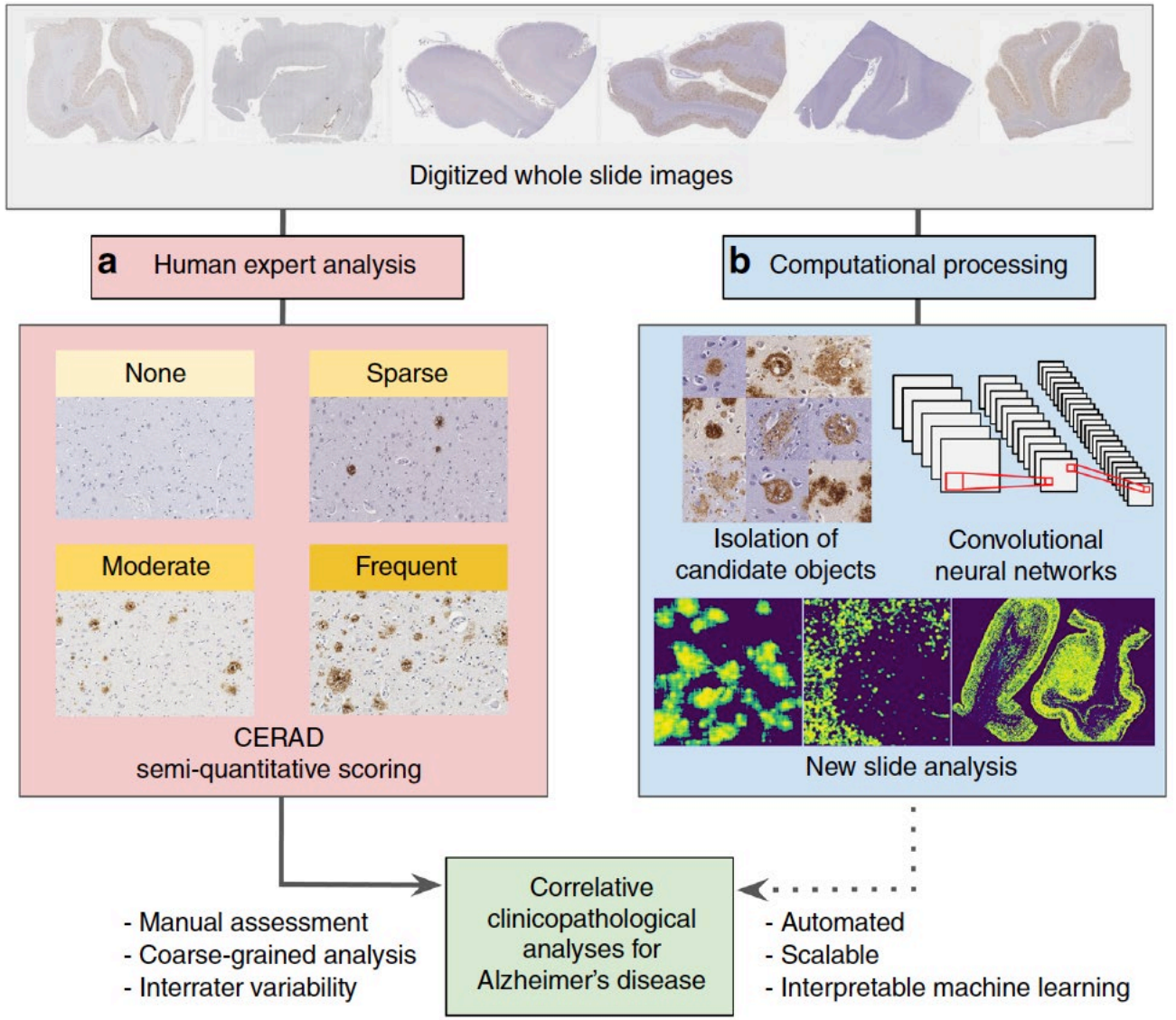
Diagnosis of AD incorporate protocols assessing **plaque density**, **distribution** and **morphology** for understanding disease progression and pathophysiology.

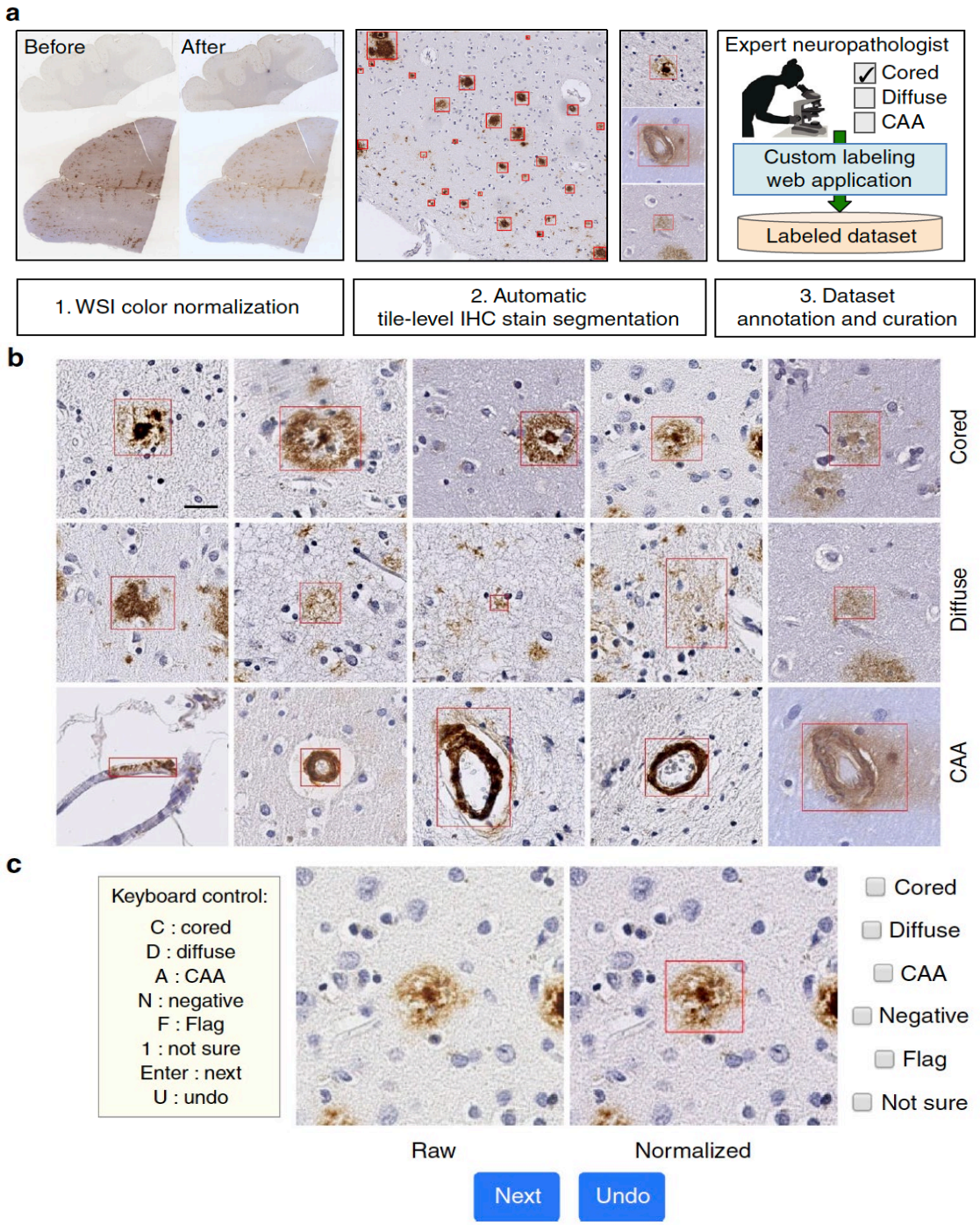
Manual counts or stereological methods can be tedious, difficult to score, and time-consuming.

DCNNs augment neuropathological whole slide image (WSI) analysis of archival human post-mortem tissues.

Custom web interface for rapid expert annotation and training of CNN models capable of distinguishing Aβ pathologies in the form of:

1. Cored plaques
2. Diffuse plaques
3. Cerebral amyloid angiopathy (CAA)


Digitized whole slide images
a Human expert analysis
b Computational processing
None
Sparse
Moderate
Frequent
CERAD
semi-quantitative scoring
Isolation of
candidate objects
Convolutional
neural networks
New slide analysis
Correlative
clinicopathological
analyses for
Alzheimer's disease
- Manual assessment
- Coarse-grained analysis
- Interrater variability
- Automated
- Scalable
- Interpretable machine learning



43 digitized microscope slides from different patients (50000x50000 pixels images)

256 X 256 pixel tiles centered on individual plaque candidates to use as direct input of the CNN

Simple web interface to rapidly annotate Aβ pathology-candidate image tiles

- 55,001 annotated images for model training have been used (85% were images of diffuse plaques).
- The trained model was applied to another dataset with 101,671 unlabeled images. These unlabeled images were ranked by the prediction confidence for cored plaques or CAAs.
- 11,029 images with high prediction confidence for cored plaques or CAAs were then labeled by the neuropathologist as annotation phase II.

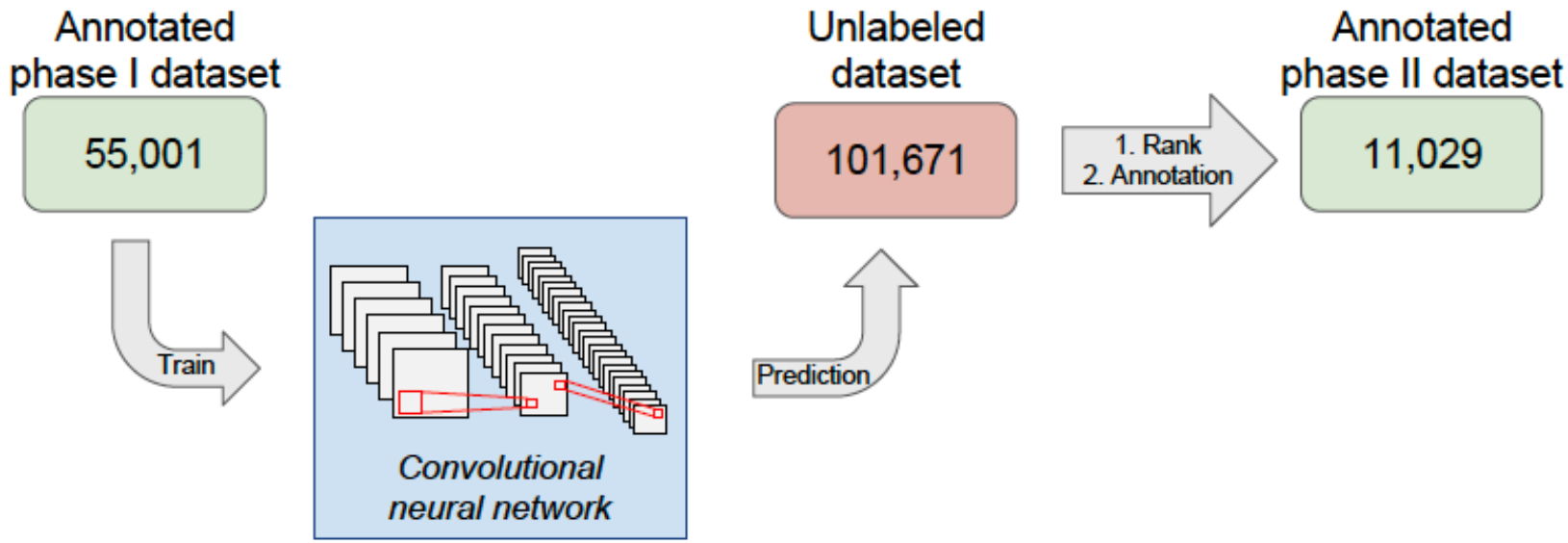


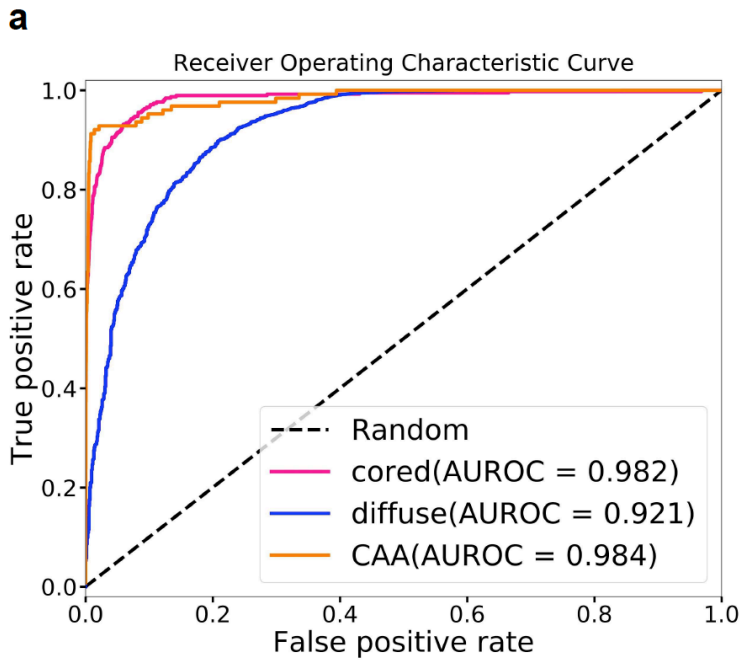
Table 1 Summary of annotated object tile dataset by project phase

Phase	Cored plaque	Diffuse plaque	CAA	Total
Development phase I	1233 (2.24%)	46,650 (84.82%)	778 (1.14%)	55,001
Development phase II	1035 (9.38%)	7610 (69.00%)	1405 (12.74%)	11,029
Development total	2268 (3.43%)	54,260 (82.17%)	2183 (3.31%)	66,030
Test (phase III)	83 (0.76%)	10,234 (94.12%)	7 (0.06%)	10,873

Remaining unlisted percentages correspond to Not Sure or Flagged labels (see Methods)

Receiver operator characteristic curve

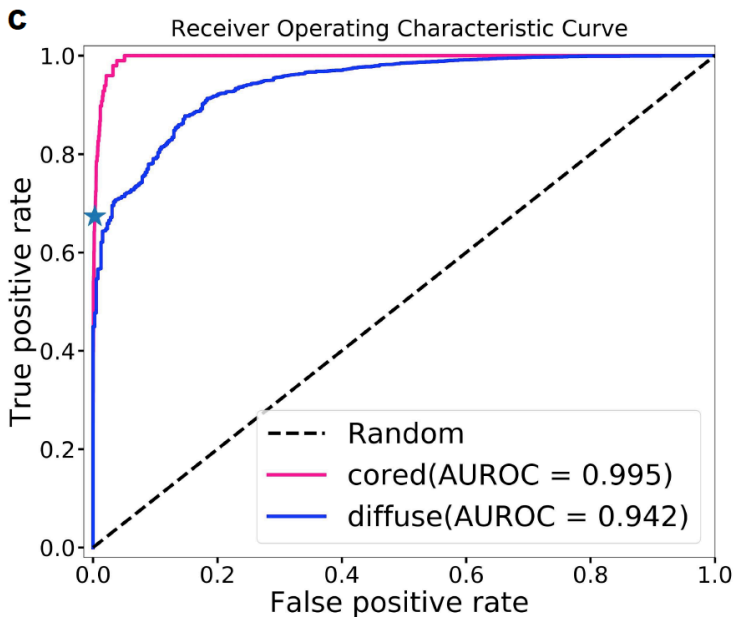
(on validation set)



$$\frac{\text{human annotated plaques pixels}}{CNN \text{ predicted plaques pixels}}$$

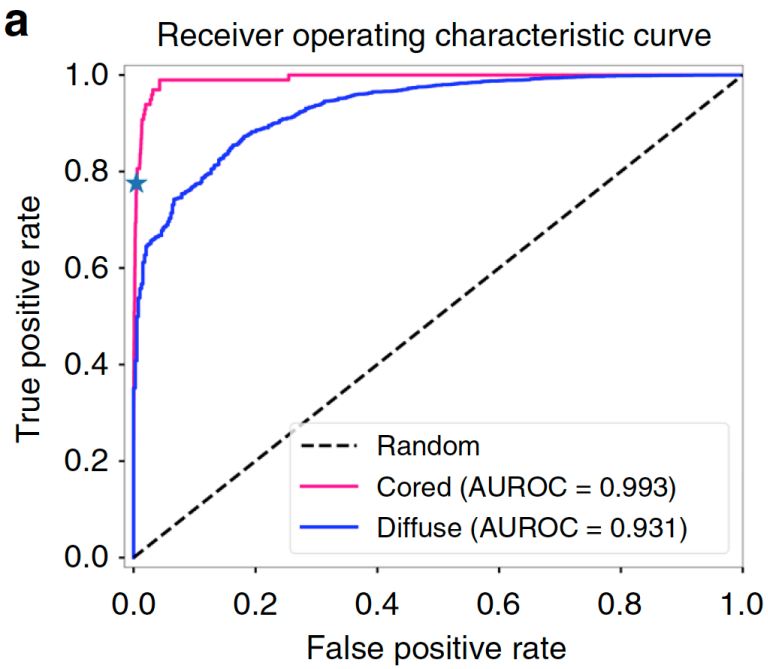
Receiver operator characteristic curve

(on test set)

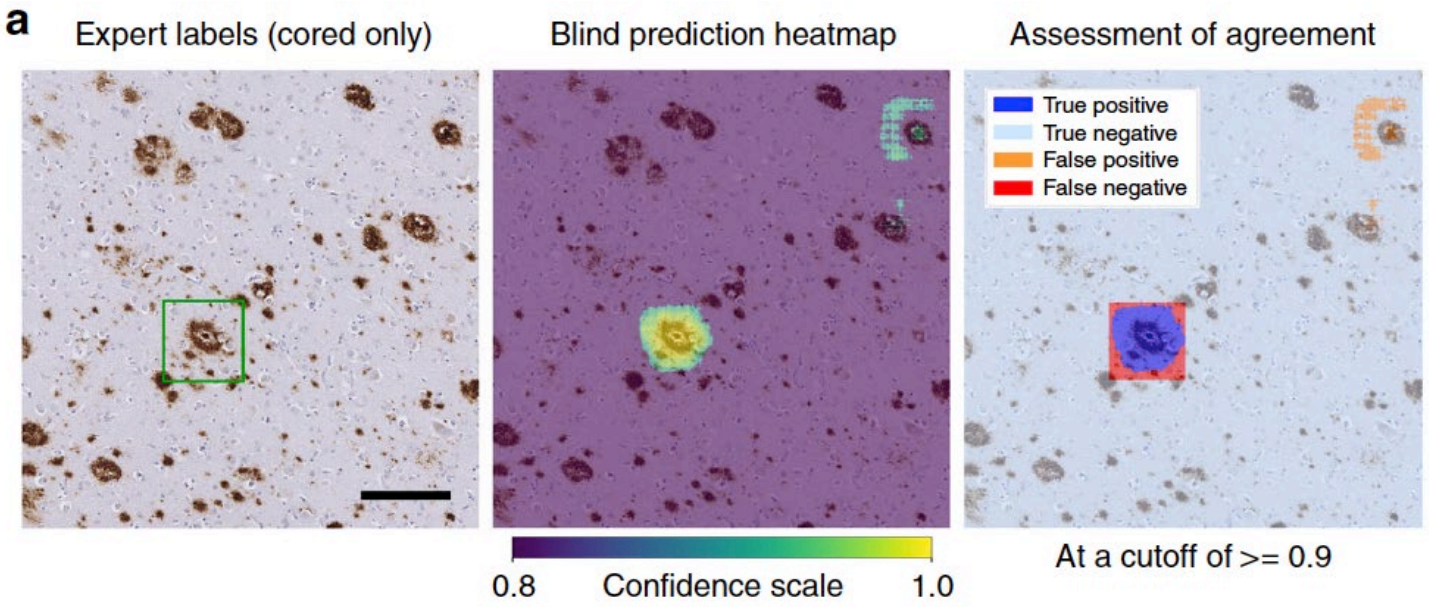


Receiver operator characteristic curve

(on test set) with 50% less annotated training images



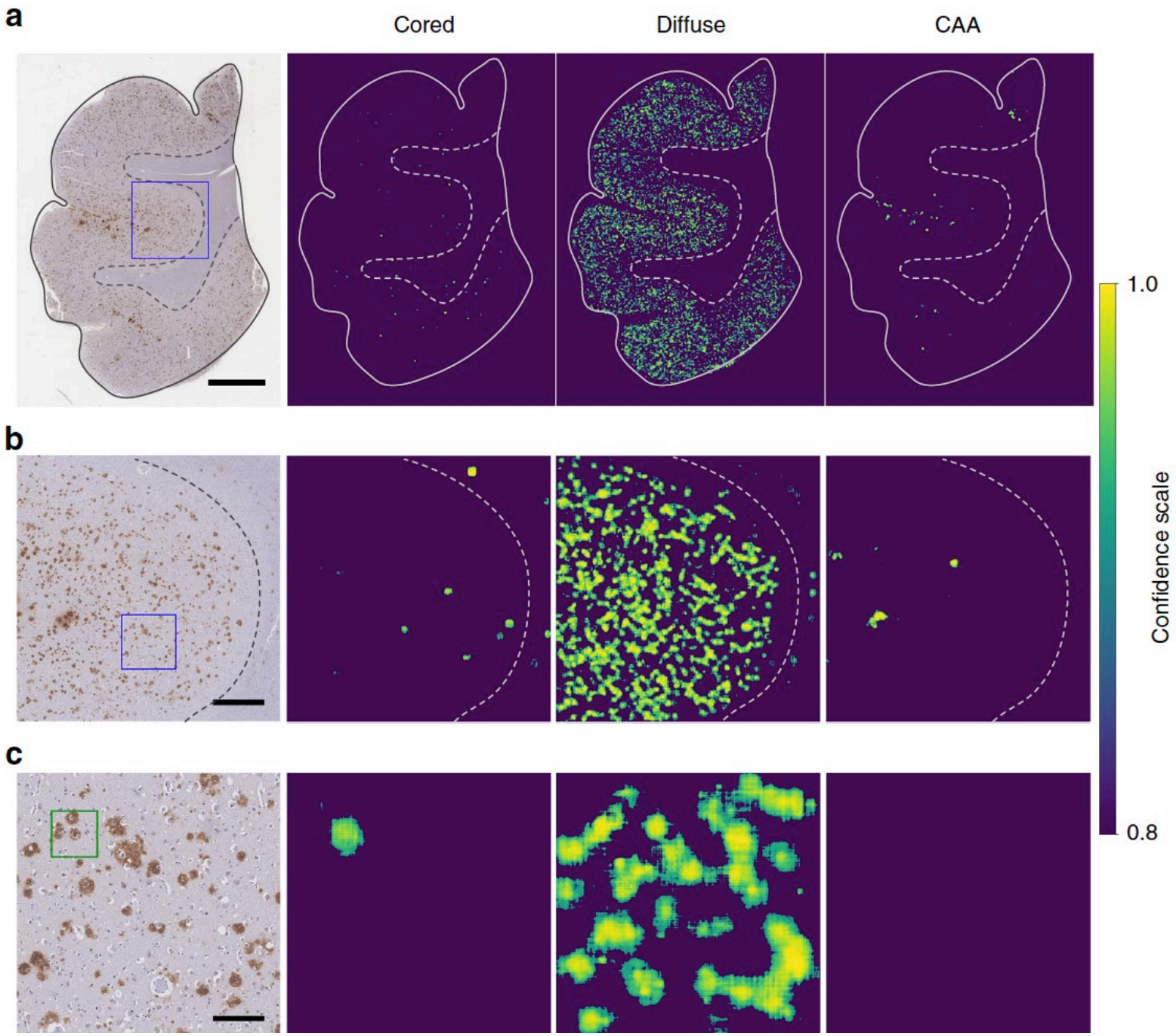
NB: CAA prediction was excluded in the test set because the 10 hold-out patients' slides lacked them.



In the **leftmost column**, a green box surrounds the cored plaque within the tile, as labeled by a neuropathologist during the Phase-III dataset annotation

The **middle column** overlays the prediction map onto the original IHC-stained image.

The **rightmost column** summarizes agreement between the expert label and the prediction, with blue and cyan representing correct prediction areas, while red and orange denote misclassification



In this cohort, diffuse plaques are densely distributed across the gray matter, whereas cored plaques are predominantly located in deeper and lower cortical layers, in accordance with known neuroanatomic distributions

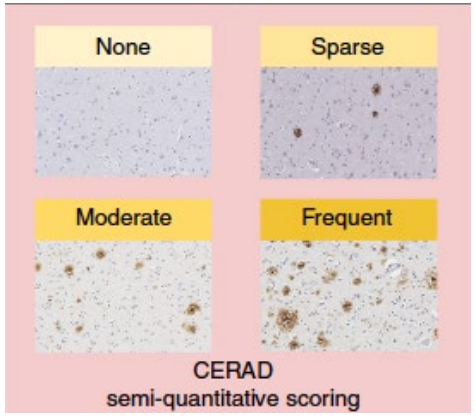
Furthermore, CAA predictions predominantly appear proximal to the cortical surface where leptomeninges are present

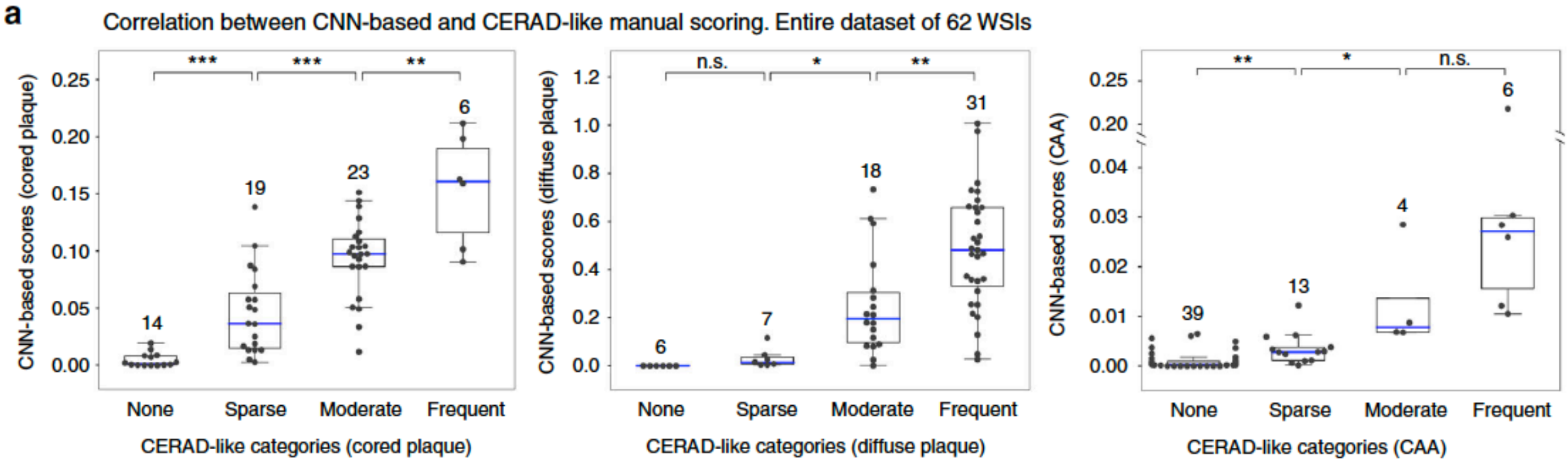
These maps highlight other locational aspects of the plaques, such as their presence in the white matter immediately beneath the gray matter.

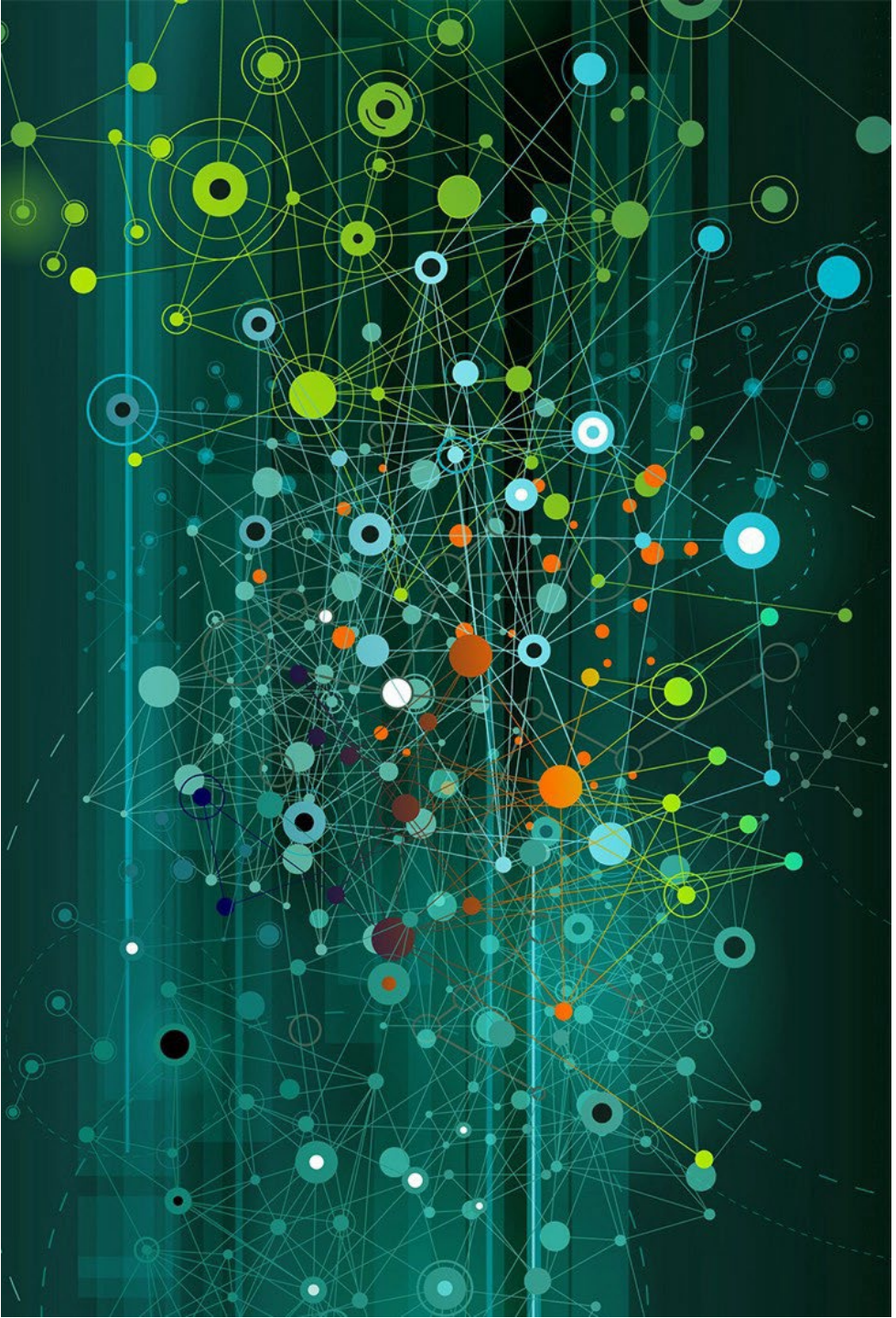
CNN-based WSI scores for Aβ pathology at a global WSI level

For the CNN-based score, they calculated a count of each predicted Aβ plaque-type across an entire WSI by segmenting its prediction heatmap and normalizing the result by the tissue area of each slide.

The resulting CNN-based scores correlated strongly across the total dataset of 62 WSIs for which their neuropathologists independently-collected semiquantitative, CERAD-like scores for each specific class on 62 of the WSIs.





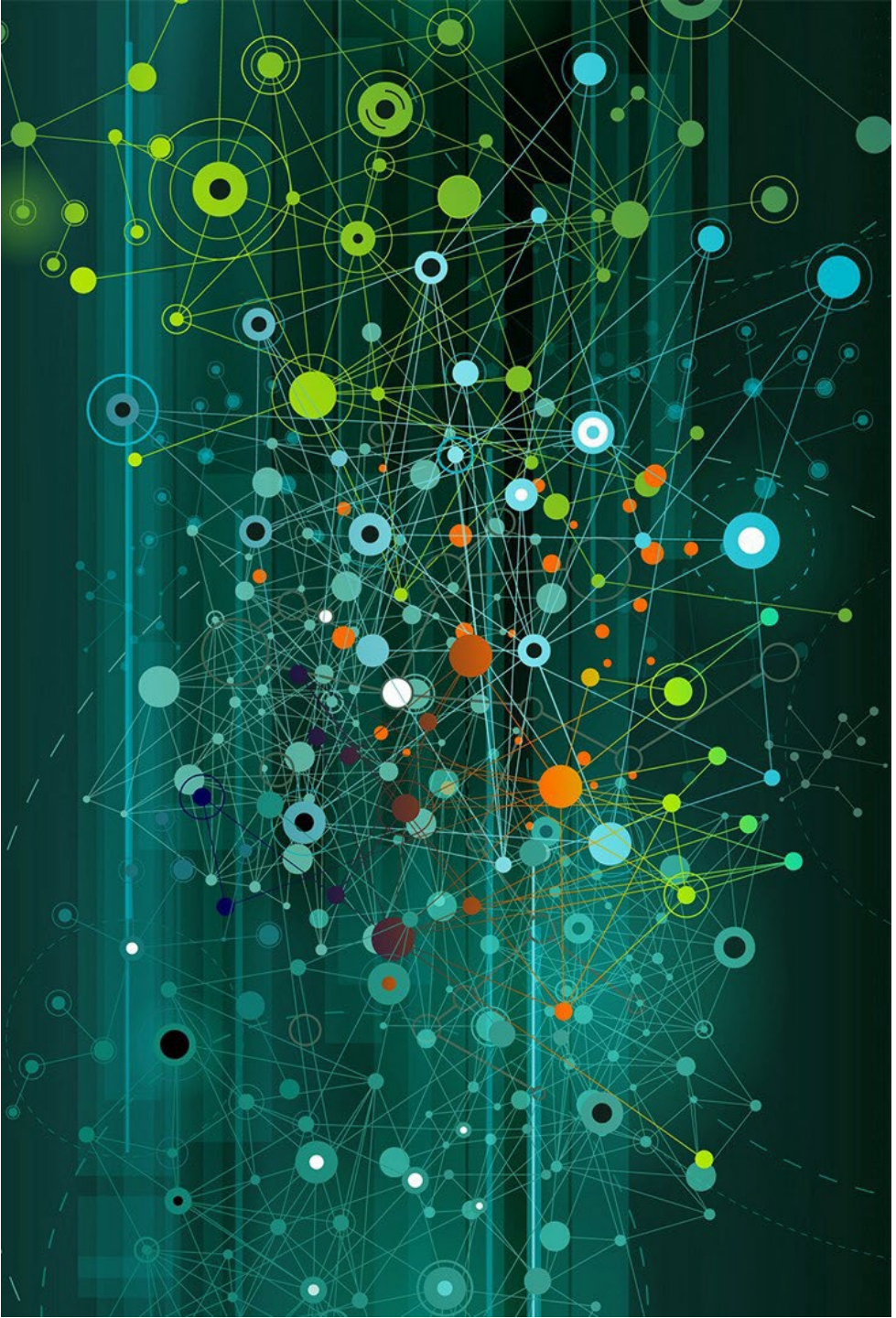


OUTLOOKS

- They developed an end-to-end pipeline to automate WSI processing and aid rapid image annotation (cloud-based Amazon Web Services Elastic Beanstalk) .

- They retrospectively evaluated the necessary training data size suggesting a reduced dataset may be pragmatically sufficient for classification of cored and diffuse plaques.

- They finally evaluated whether **CNN models could automatically quantify Aβ burden** on a whole-slide level in a way that would correlate with standard semi-quantitative methods for plaque assessment.



PROS

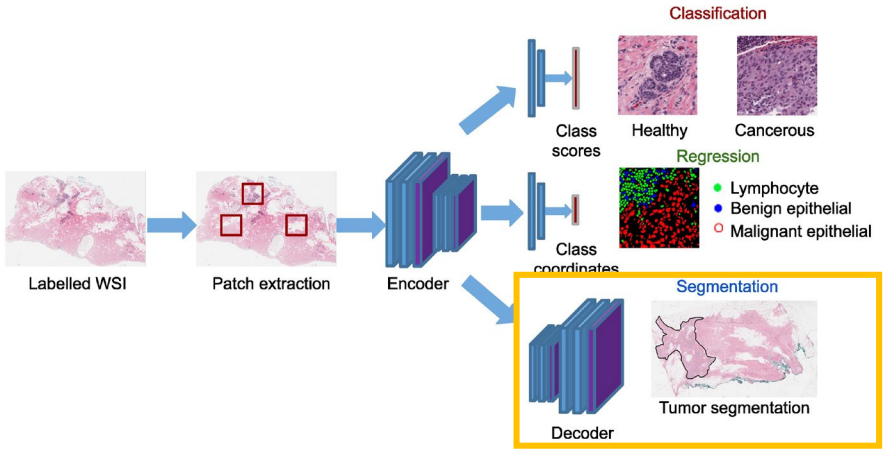
- Models such as these may help to quantify contents of pathologies in a scalable way.

- It may reduce the pathologists' workload giving a second opinion on multiple pathological specific phenotypes

CONS

- Differences in experience and annotation criteria will likely result in individual expert variation among ground truth labels.

- All data used in this study were from a single brain bank and retrieved and digitized under the same conditions; more diverse datasets from multiple sources will yield more robust and reliable models.

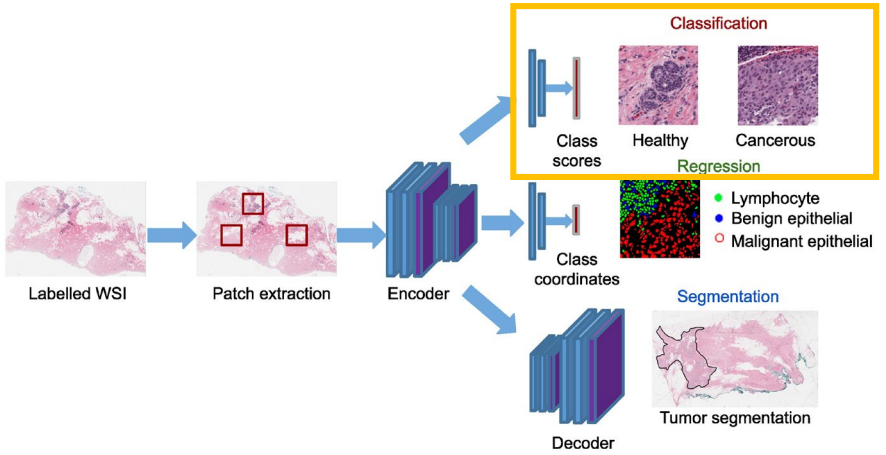


nature COMMUNICATIONS

https://doi.org/10.1038/s41467-019-10212-1 OPEN

Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline

Ziqi Tang[1,2], Kangway V. Chuang[1], Charles DeCarli[3], Lee-Way Jin[4], Laurel Beckett[5], Michael J. Keiser[1] & Brittany N. Dugger[6]



ARTICLES
https://doi.org/10.1038/s42256-019-0052-1
nature machine intelligence

Corrected: Publisher Correction; Publisher Correction

Pathologist-level interpretable whole-slide cancer diagnosis with deep learning

Zizhao Zhang[1], Pingjun Chen[2], Mason McGough[2], Fuyong Xing[3], Chunbao Wang[4], Marilyn Bui[5], Yuanpu Xie[2], Manish Sapkota[6], Lei Cui[2], Jasreman Dhillon[5], Nazeel Ahmad[7], Farah K. Khalil[5], Shohreh I. Dickinson[5], Xiaoshuang Shi[2], Fujun Liu[6], Hai Su[2], Jinzheng Cai[2] and Lin Yang[2*]

Supervised Learning – Classification models

Within this category, we further identify two sub-categories:

- **Local-level tasks**: based on a region (i.e., cell, nuclei) represented by a spatially pooled feature representations or scores, aiming at identifying or localizing objects.

- **Global-level tasks**: related to image-level prediction tasks such as whole-slide level disease grading (e.g. tissue-level cancer localization)

The main success of these CNN models depends on:

- The number of images available for training
- The choice of network hyper-parameters
- Various other boosting techniques

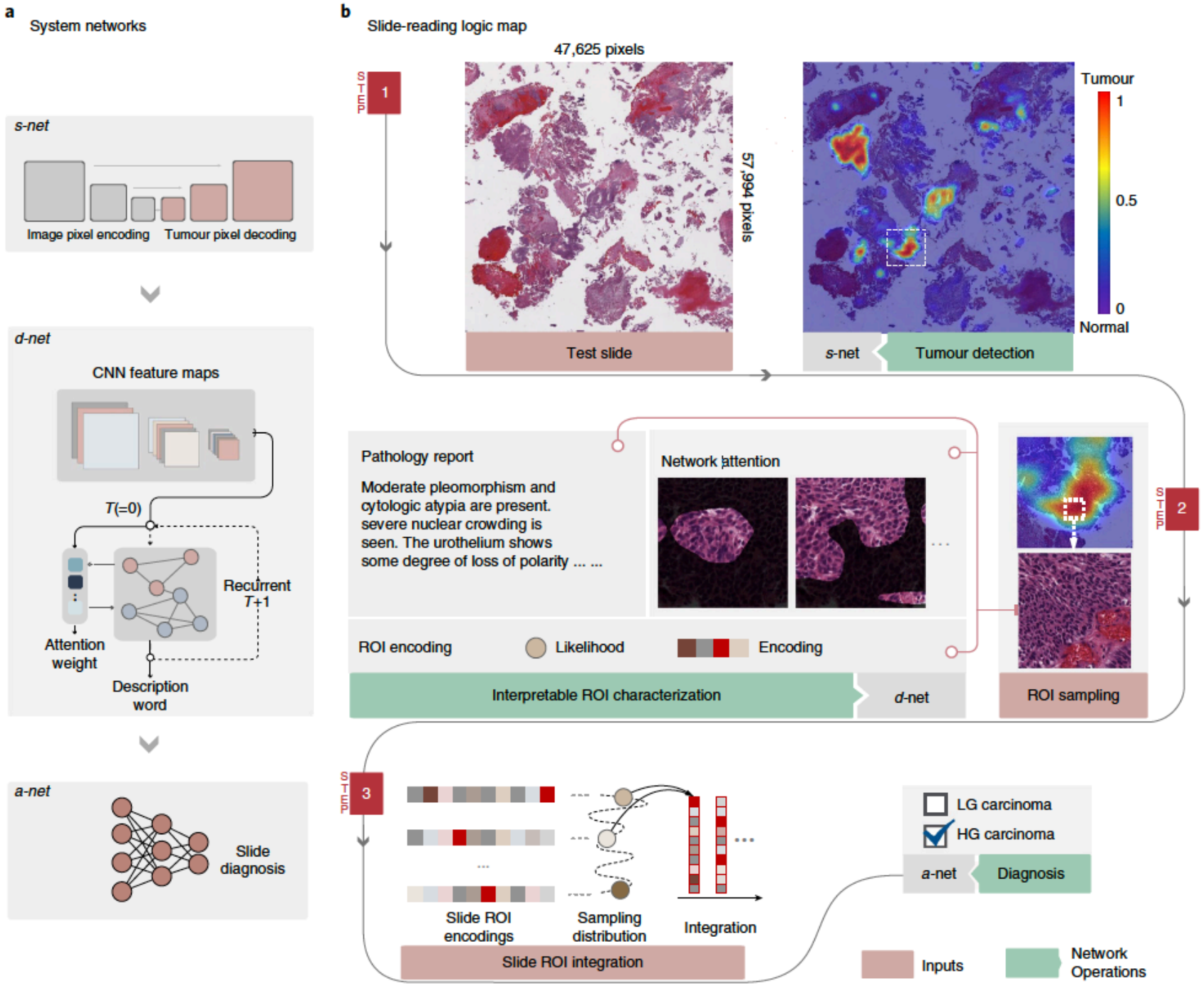
ARTICLES
https://doi.org/10.1038/s42256-019-0052-1
nature machine intelligence
Corrected: Publisher Correction; Publisher Correction

Pathologist-level interpretable whole-slide cancer diagnosis with deep learning

Zizhao Zhang[1], Pingjun Chen[2], Mason McGough[2], Fuyong Xing[3], Chunbao Wang[4], Marilyn Bui[5], Yuanpu Xie[2], Manish Sapkota[6], Lei Cui[2], Jasreman Dhillon[5], Nazeel Ahmad[7], Farah K. Khalil[5], Shohreh I. Dickinson[5], Xiaoshuang Shi[2], Fujun Liu[6], Hai Su[2], Jinzheng Cai[2] and Lin Yang[2*]

Intro & Goals

- AI to generate clinical diagnostic descriptions and network visual attention maps of whole-slide data representing patients with **bladder cancer**.
- A novel pathology whole-slide diagnosis method, powered by artificial intelligence, to address the lack of interpretable diagnosis.
- It automates the human-like diagnostic reasoning process and translate gigapixels directly to a series of interpretable predictions
- It provides a second opinions and thereby encouraging consensus in clinics.



Validated it on a large dataset containing 913 haematoxylin and eosin (H&E) stained whole slides from patients with bladder cancer.

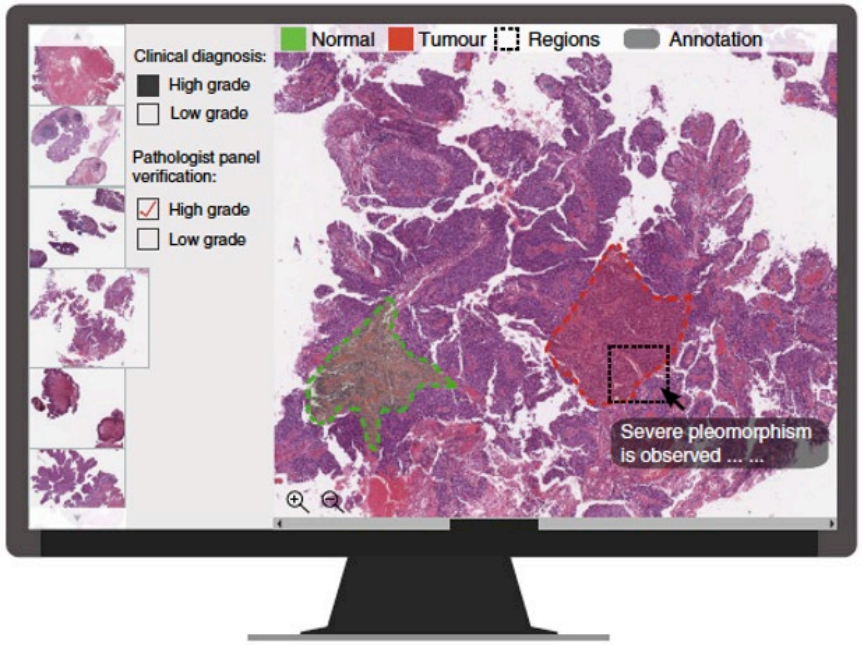
21 pathologists were involved:

- N=4 in segmenting the training images
- N=17 in performance evaluations

s-net (first network)

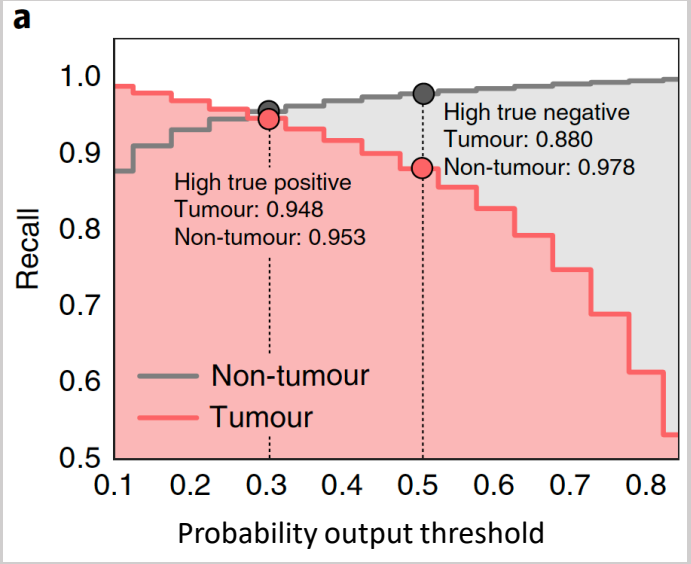
The s-net conducts tumour detection by classifying each pixel as tumour or non-tumour, represented as a **probability value**.

The resulting map indicates a diagnostically useful area, allowing the rapid localization of tumour regions

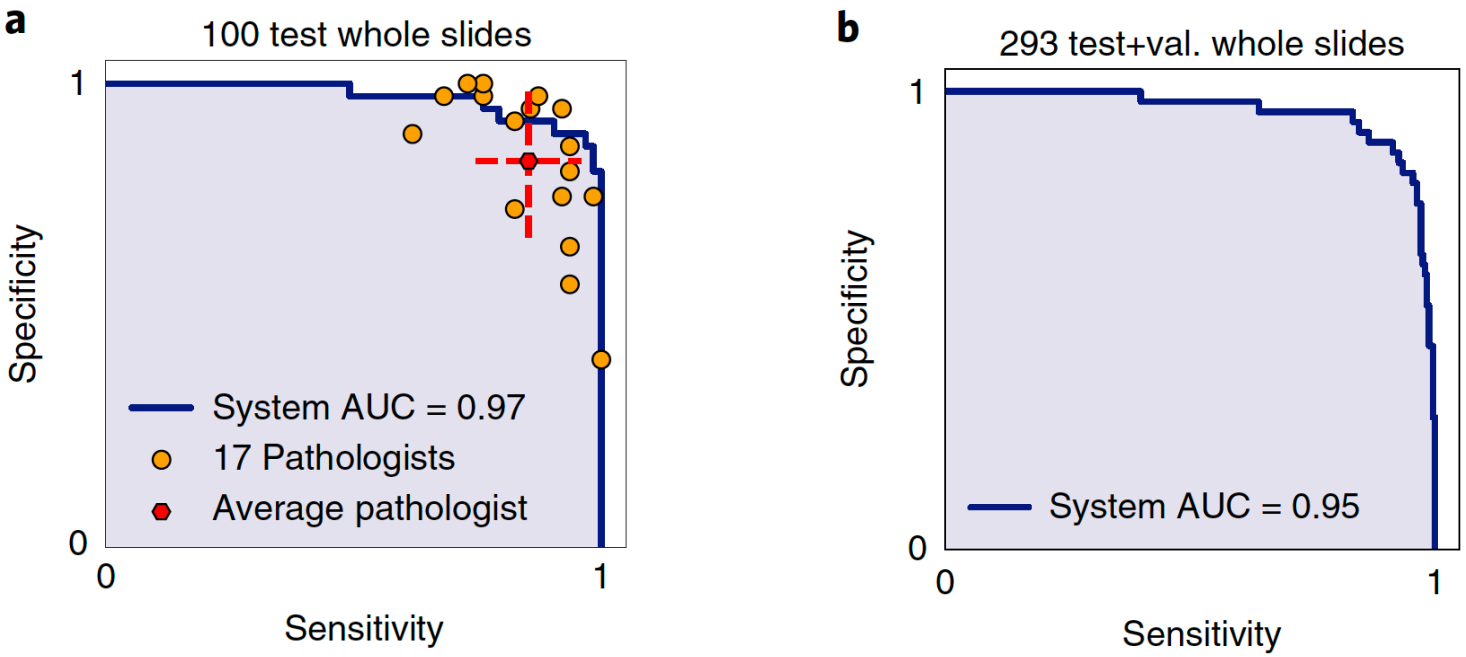


What proportion of actual positives was identified correctly?

$$Recall = \frac{TP}{TP + FN}$$



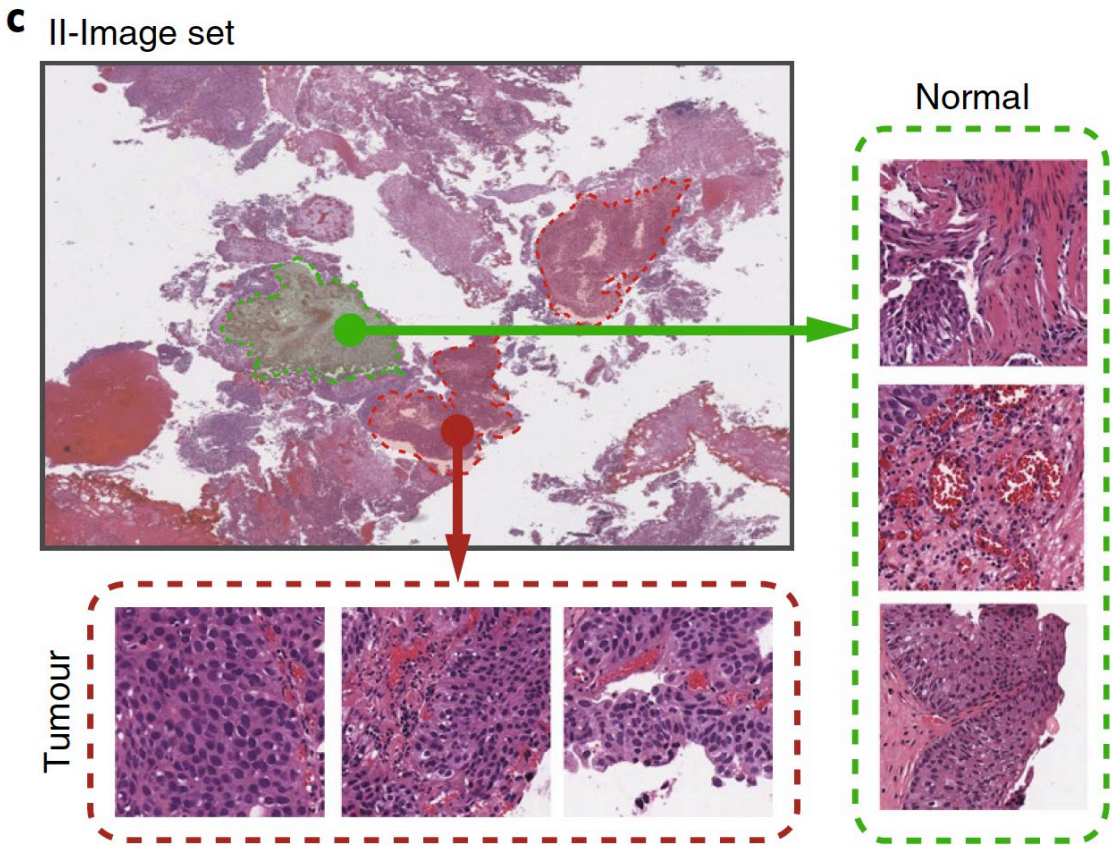
s-net (first network)



Dataset summary

ID	Type	(Data, annotation)	Train	Validation	Test
I	Slide	(Slide, mask&label)	620	193	100
II	Image	(Image, mask&label)	148,671	8,371	–
III	Report	(Image, text)	11,820	6,297	3,148
IV	Diagnosis	(Feature, label)	620·*M*	193·*M*	100·*M*

s-net (first network)

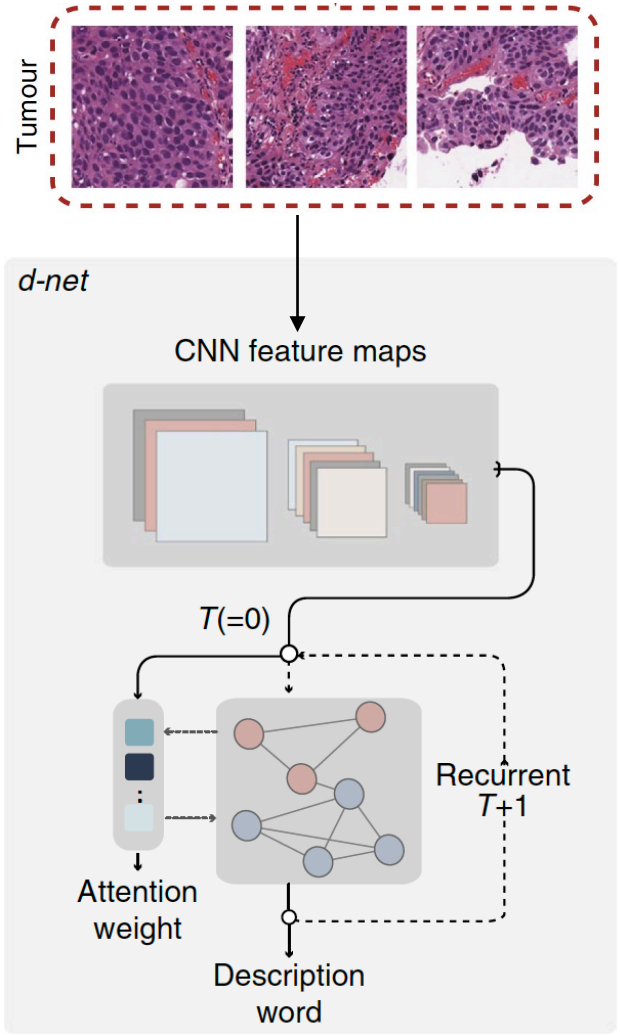


Dataset summary

ID	Type	(Data, annotation)	Train	Validation	Test
I	Slide	(Slide, mask&label)	620	193	100
II	Image	(Image, mask&label)	148,671	8,371	–
III	Report	(Image, text)	11,820	6,297	3,148
IV	Diagnosis	(Feature, label)	620·*M*	193·*M*	100·*M*

d-net (second network)

Based on the tumour detection result, the system conducts the following steps to automatically select a set of diagnostically useful tissue images around detected tumours from the whole slide and in turn associate the pathology reports containing five morphological features and the final diagnose.



The d-net is a composite neural network that can combine multimodal information → It includes:

Inception-v3 CNN → an image model to represent visual knowledge by encoding image pixels into feature maps.

advanced RNN LSTM → a language model to generate diagnostic descriptions and network visual attention.

d-net (second network)

To construct the III-Report dataset, they selected 221 non-invasive HG and LG papillary urothelial carcinoma slides

↓

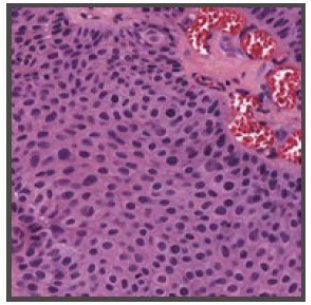
4,253 cropped images of the most representative regions

↓

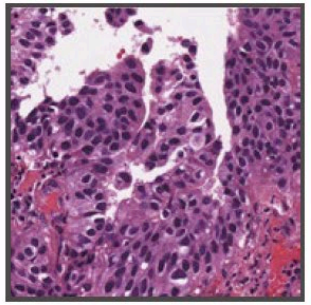
Experienced pathologists provided pathology reports containing 5 types of known pathology cellular features (state of nuclear pleomorphism, cell crowding, cell polarity, mitosis and prominence of nucleoli)

↓

21,265 image-report pairs in total



Marked variability in nuclear size shape and outline consistent with severe pleomorphism. There is a moderate degree of crowding ... ...



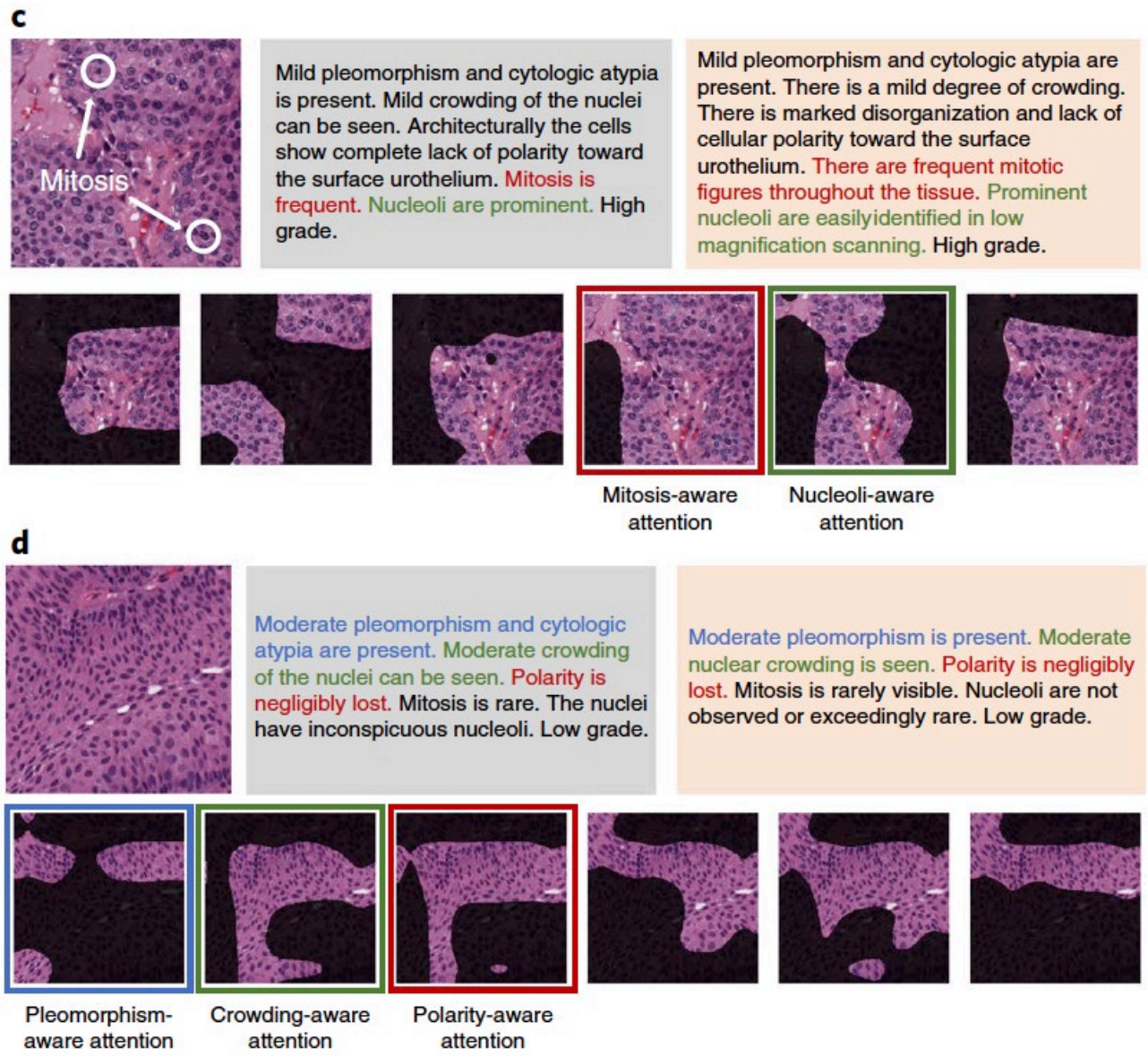
... ... Cells show complete lack of polarity. Mitosis is rare. The nuclei have inconspicuous nucleoli. High grade.

Dataset summary

ID	Type	(Data, annotation)	Train	Validation	Test
I	Slide	(Slide, mask&label)	620	193	100
II	Image	(Image, mask&label)	148,671	8,371	–
III	Report	(Image, text)	11,820	6,297	3,148
IV	Diagnosis	(Feature, label)	620·*M*	193·*M*	100·*M*

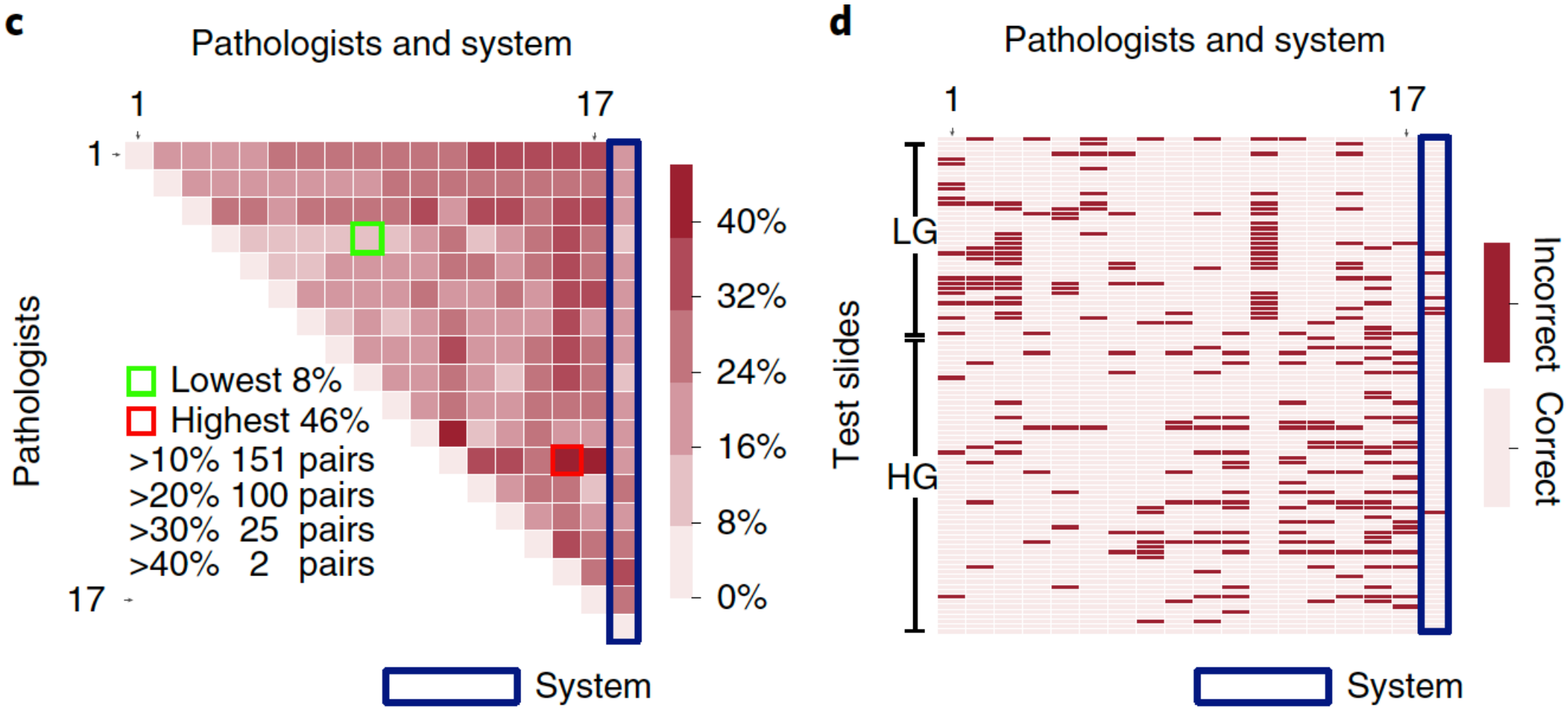
d-net (second network)

Their method accurately describes multiple types of cell features that resemble the pathologists' interpretations.



d-net (second network)

They also investigated **pathologist variability** in diagnosing bladder cancer and further compared the results to our method.



a-net (third network)

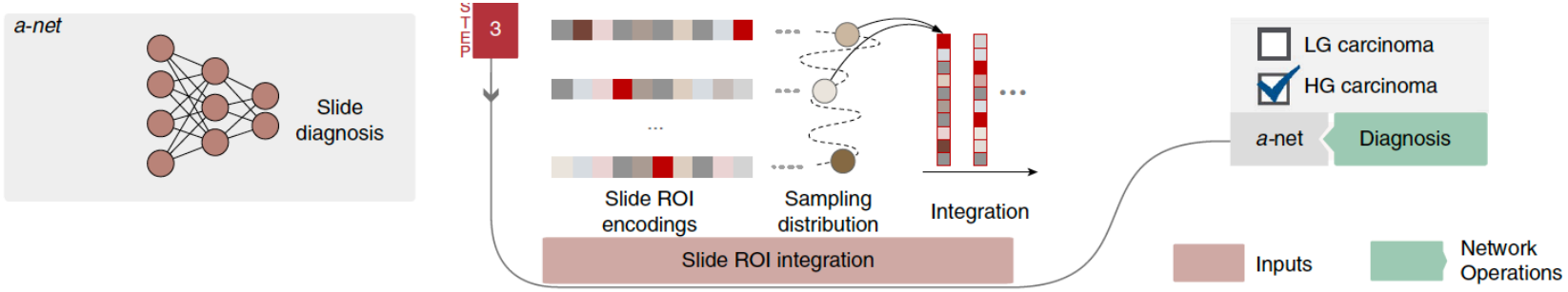
It aggregates all the diagnostic information together in all slide ROIs and establish a final diagnosis

[d-net encoded features and raw tumour class probability]

It takes integrated ROI feature encodings and predicts slide cancer labels.

a-net was trained:

- Some LG-ROIs from the d-net have been annotated as LG cancer
- Some HG-ROIs from the d-net have been annotated as HG cancer

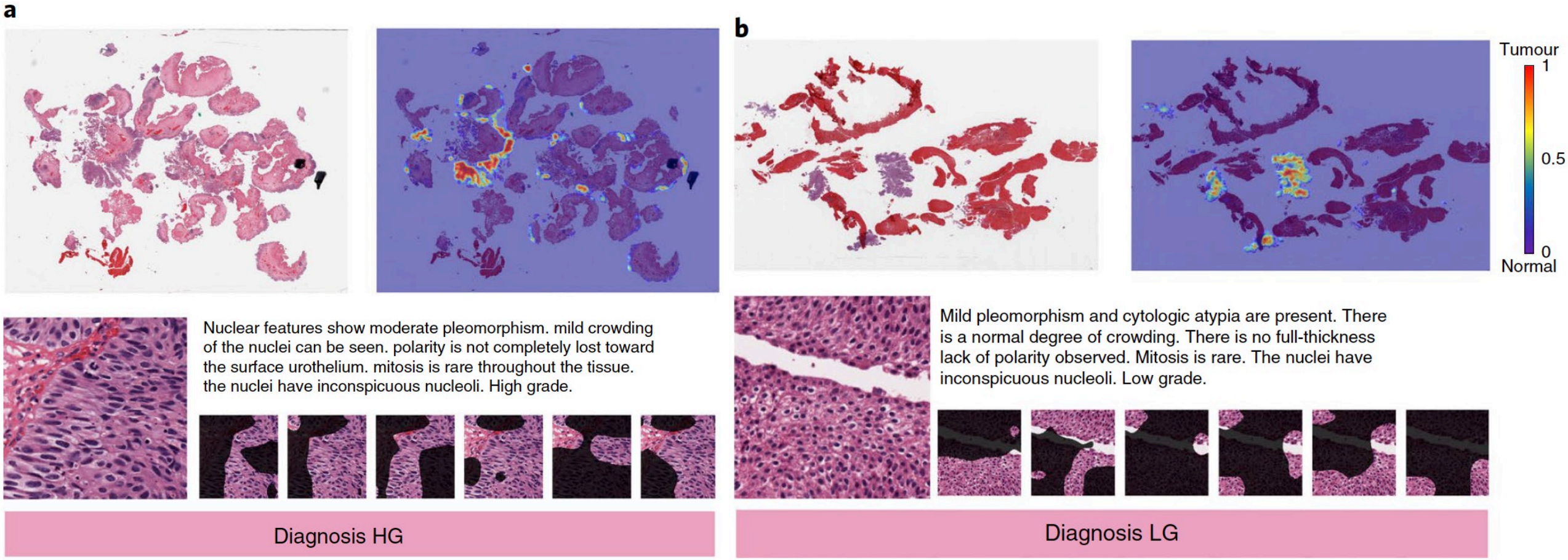


The final diagnosis is the maximum class probability response of the accumulated probability of the 10-time predictions.

They propose a stochastic feature sampling mechanism to effectively augment training data through **random feature combination** so as to improve the model generalization.

a-net (third network)

It aggregates all the diagnostic information together in all slide ROIs and establish a final diagnosis



OUTLOOK

- Novel interpretable diagnosis method for diagnostic pathology which shows unprecedented advantages over previous work.

- Comprehensive validation on a large-scale bladder cancer data set demonstrates that the performance is similar to that of a wide range of pathologists.

- It supports pathologists when conducting a second review and visual inspection.

PROS

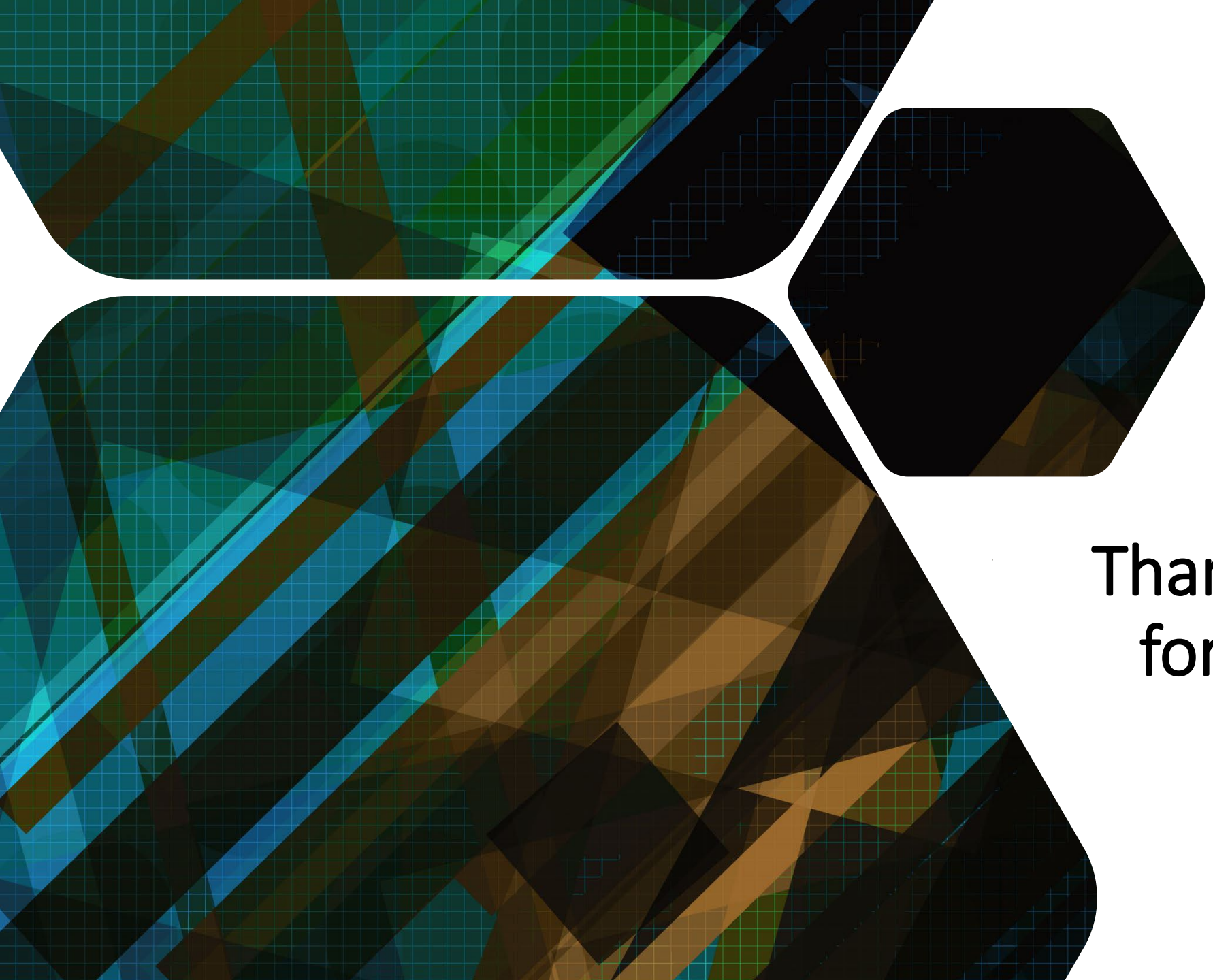
- Their method has strong generalizability for learning complex tissue structures and cell patterns in different pathological conditions.

- It may reduce the pathologists' workload giving a second opinion on multiple pathological specific phenotypes

CONS

- It is the relatively long computational time required to carry out a dense patch- wise prediction over an entire WSI.

- It requires a large number of images for training, choice of network hyper-parameters and various other boosting techniques (time-consuming to set this up)



Thank you so much
for your attention