Leadership Development (LEAD) Workshop at NAMS
09.03.2024

AI in Medicine:

Automated Detection of Gallbladder and Breast Cancer

Chetan Arora

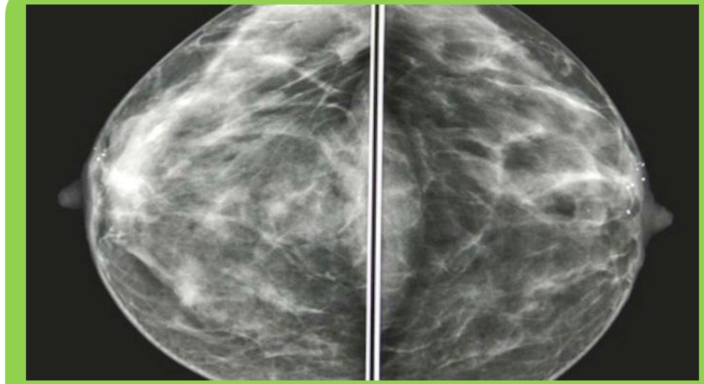
Professor, Computer Science and Engineering. Joint Faculty School of AI

Indian Institute of Technology Delhi

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

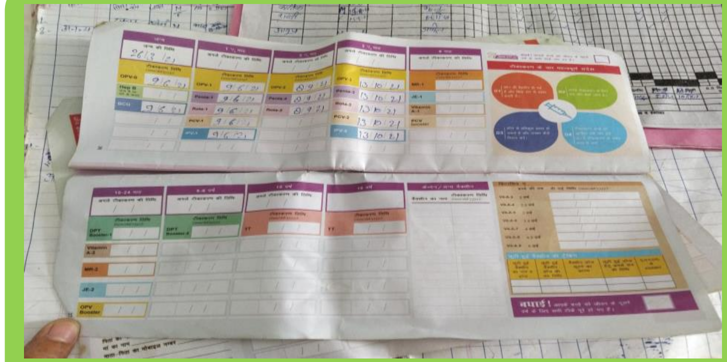
Our Work (in digital healthcare)

Gall Bladder Cancer Detection



Breast Cancer Detection

Neurosurgery Training and Eval.

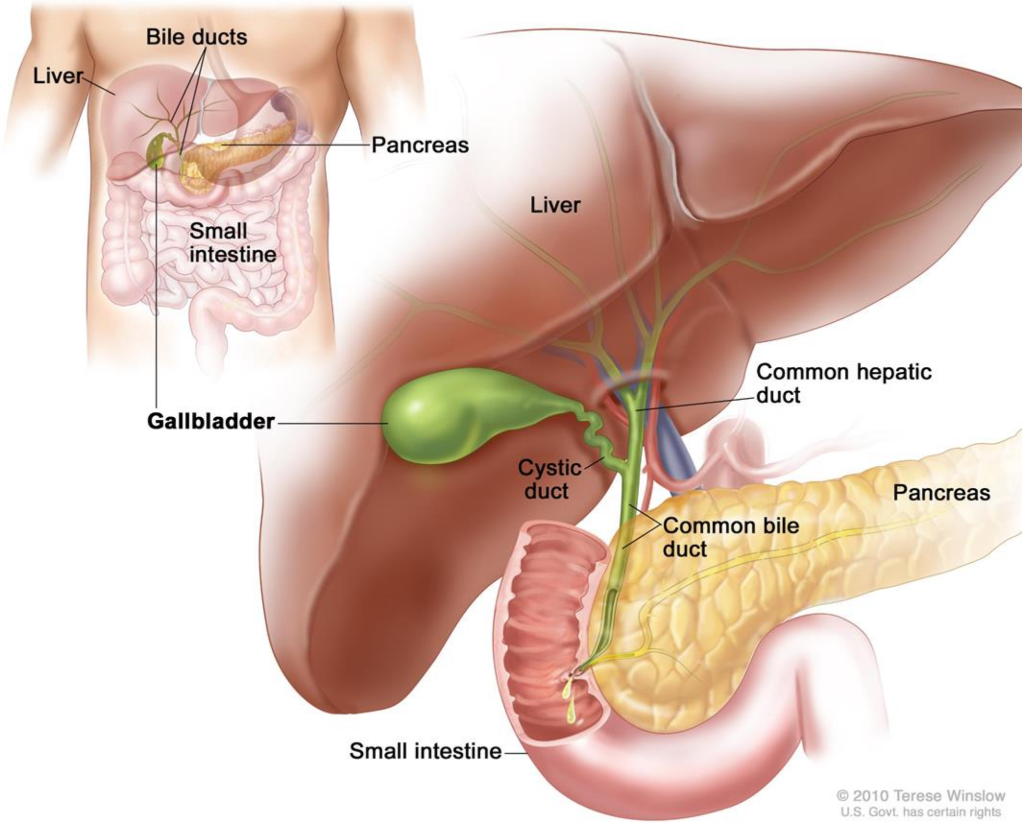


Public Health Data Management

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Gallbladder Cancer (GBC)

- Worldwide - 84,000 deaths every year [1]
- 5-year survival rate is 5%. Mean survival - 6 months
- Quick Metastasis - contiguous liver tissues
- Silent progress - often detected at a very late stage



[1] GLOBOCAN 2020

Motivation

- Early detection and resection can increase 5-year survival rate to 44%. [1]
- Non-expert radiologists perform poorly on US images, even expert radiologist has about ~70% sensitivity for GBC detection
- AI-based automated detection for second reading - improve accuracy, triage

[1] C. Chen et al. Long-Term Outcomes and Prognostic Factors in Advanced Gallbladder Cancer: Focus on the Advanced T Stage. PLOS ONE 2016, https://doi.org/10.1371/journal.pone.0166361

Why Ultrasound Sonography (USG) for GBC?

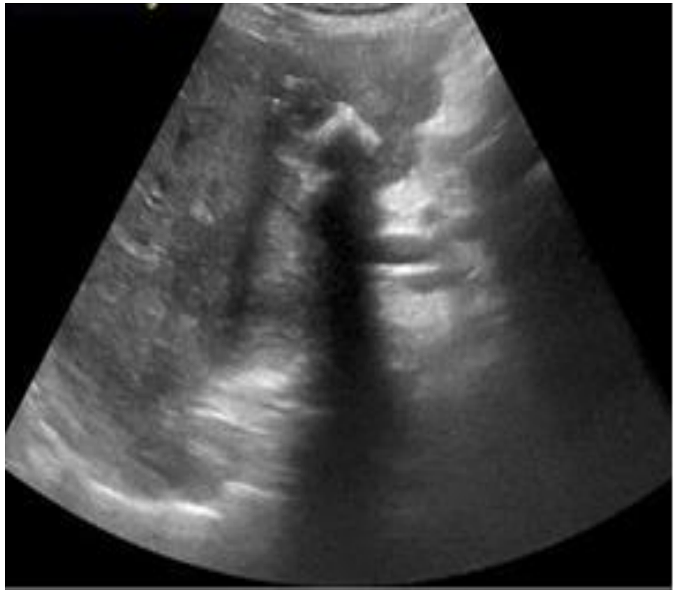
- Most common imaging modality for abdominal ailments – often the sole diagnostic imaging performed
- Highly accessible and low cost (compared to CT/ MRI)
- Excellent candidate modality for GBC detection
- No work on AI/ML-based GBC detection from US prior to ours

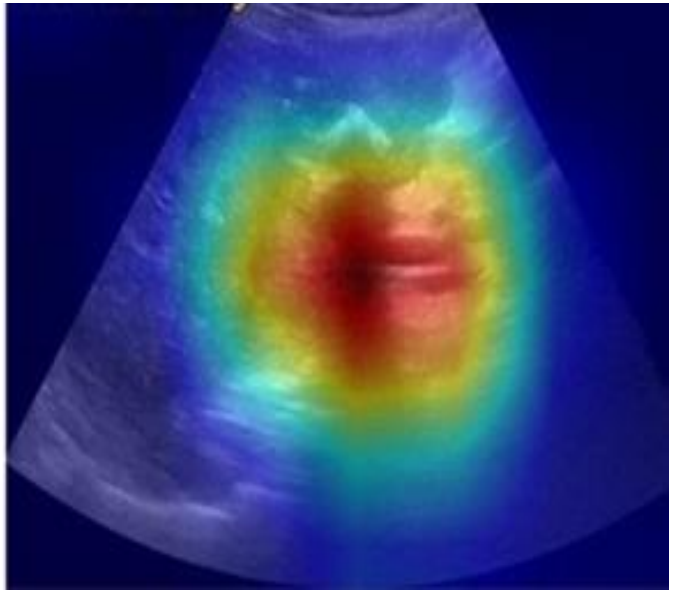
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Challenges with USG Modality

- **Low Image Quality**
  - Noise, artifacts such as shadow, and spurious textures

- **Handheld Sensor**
  - Hand-held - high degree of variability across radiologists, and medical centers.



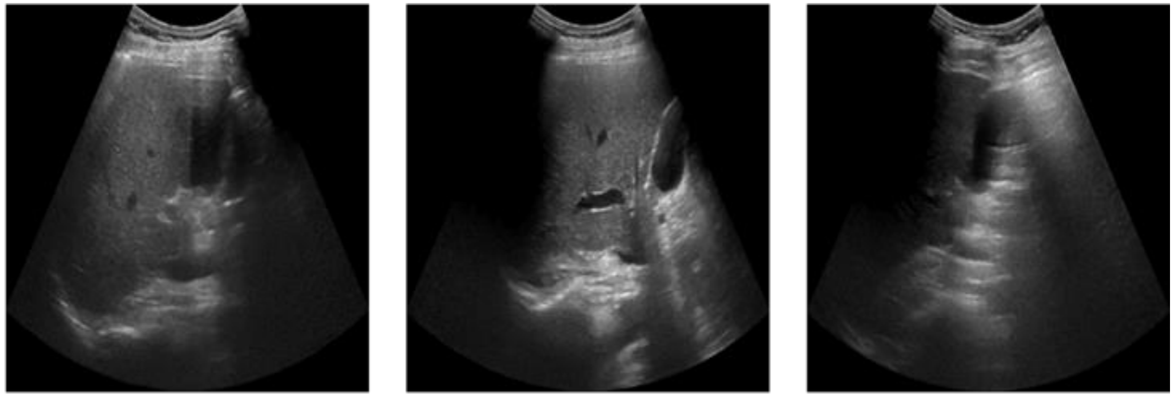


Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Challenges with USG Modality

- **Anatomy**
  - Non-regular anatomy of malignant gallbladder (loss of interface with adjacent organs, irregular anatomical structure)
  - Low inter-class variance, High intra-class variance

Low inter-class variability



High intra-class variance

- **Lack of Annotated Dataset**

Key Research Questions

- How to tackle challenges posed by Ultrasound images to make accurate predictions?
  - Artefacts such as noise, shadow, spurious textures – low image quality
  - Handheld sensor – variability in viewpoints

- Can we add interpretability?
  - Interpretable decision making by the models

Key Research Questions

- Can we learn from limited supervised data?
  - Specialized annotation – scarce labelled data
- How to design reliable models – trustworthy predictions?
- How does the AI models perform to different GBC patient subgroups?

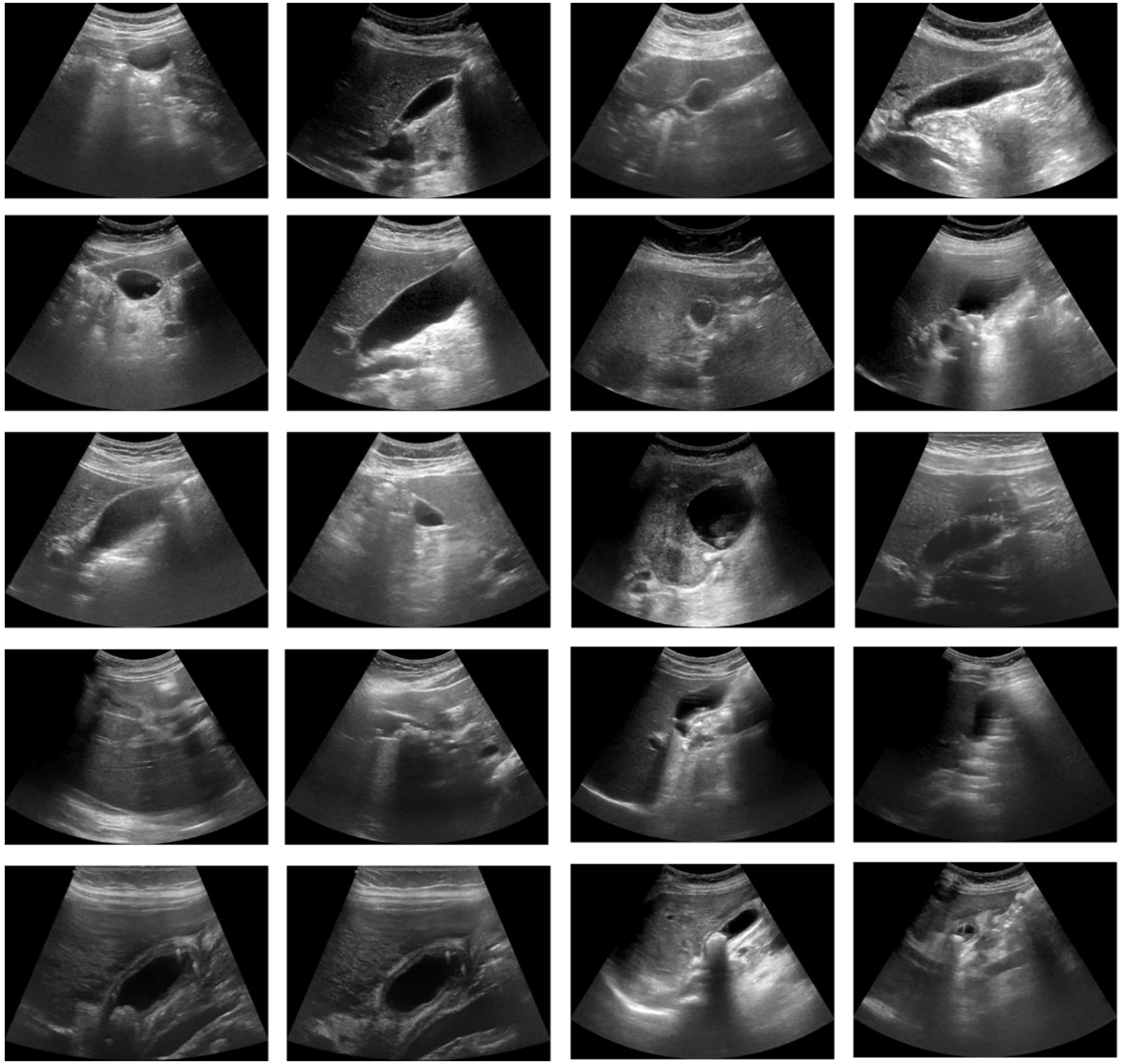
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

GBCU Dataset

- We contribute first public dataset (GBCU Dataset) for detecting GBC from USG images

- 1255 samples from 218 patients
- 990 non-malignant, and 265 malignant image samples
- Biopsy-proven ground truth

https://gbc-iitd.github.io/data/gbcu

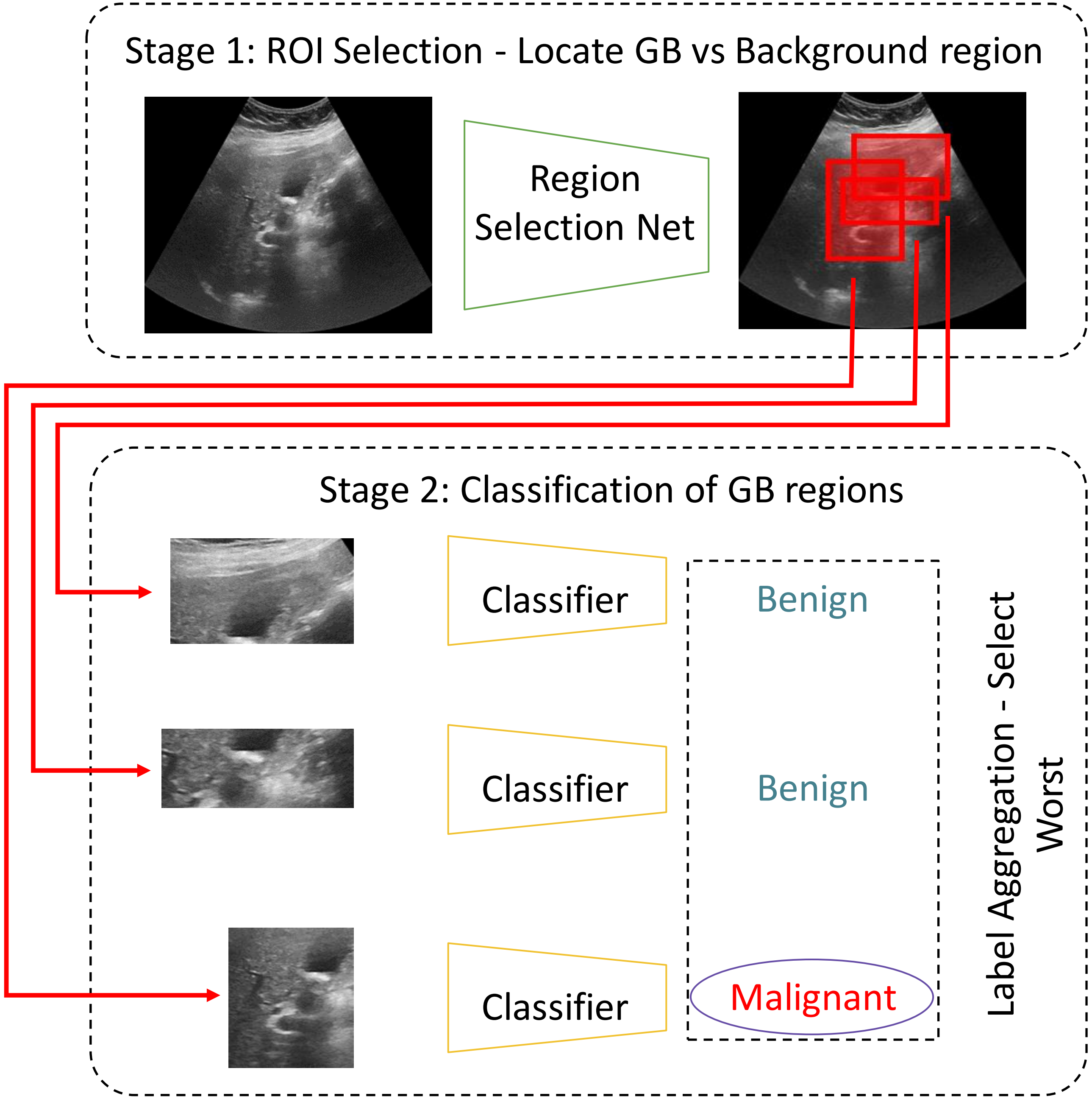


Designing Accurate GBC Detectors
(CVPR 2022)

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Our Solution: GBCNet

- Focused attention regions (ROI)
    - Reduces effect of artifacts
- Multi-scale second order pooling (MS-SoP) classifier
    - Capture different appearances
    - Rich features for malignancy



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Visual Acuity

- Humans start visual experience with blurred vision (low VA) in infancy and then gain clarity (high VA) as they grow
- Blurry images do not contain enough local information like textures - the cortex tries to expand the receptive field to capture global features such as shape of objects

Low VA - not enough local information

Expand RF for more information like shape

High VA - distinguish using local information

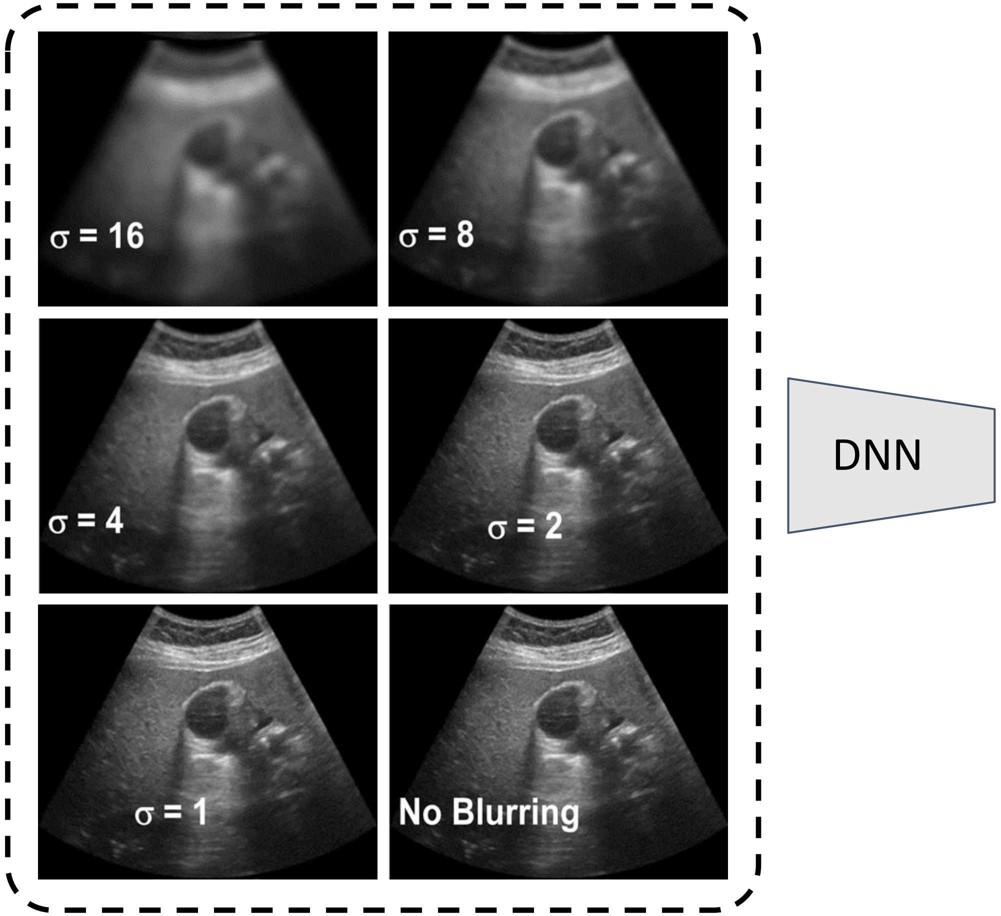
Also use expanded view learned in infancy

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Visual Acuity Curriculum - Tackle Texture Bias

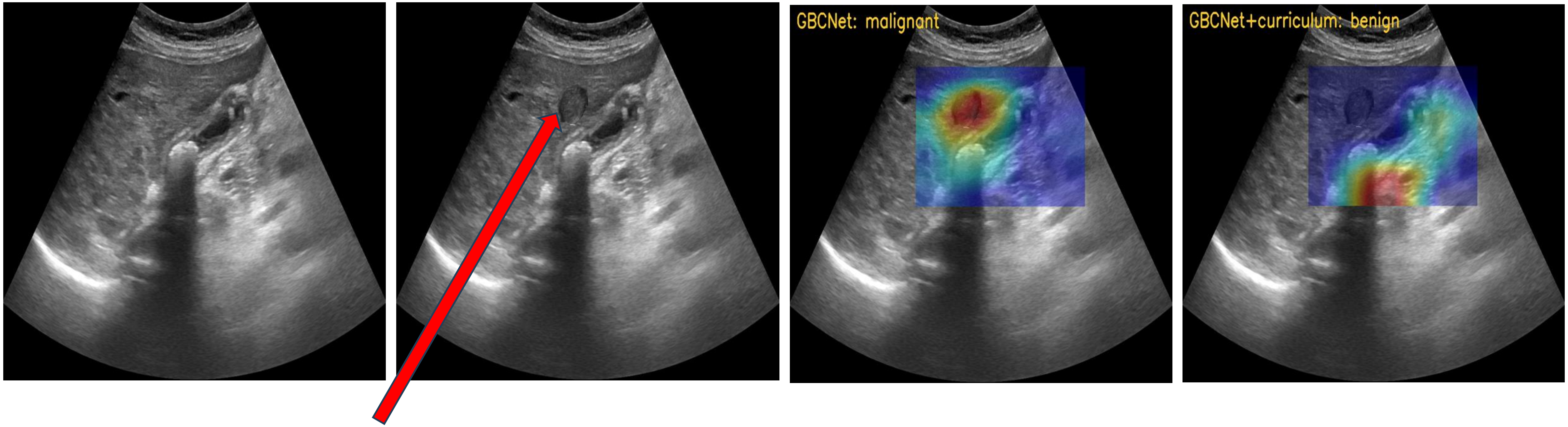
- Gaussian Blurring to simulate visual acuity and used as a training curriculum

- Start training with blurred images - gradually introduce high resolution images



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Curriculum in Tackling Textures



Synthetic
Texture

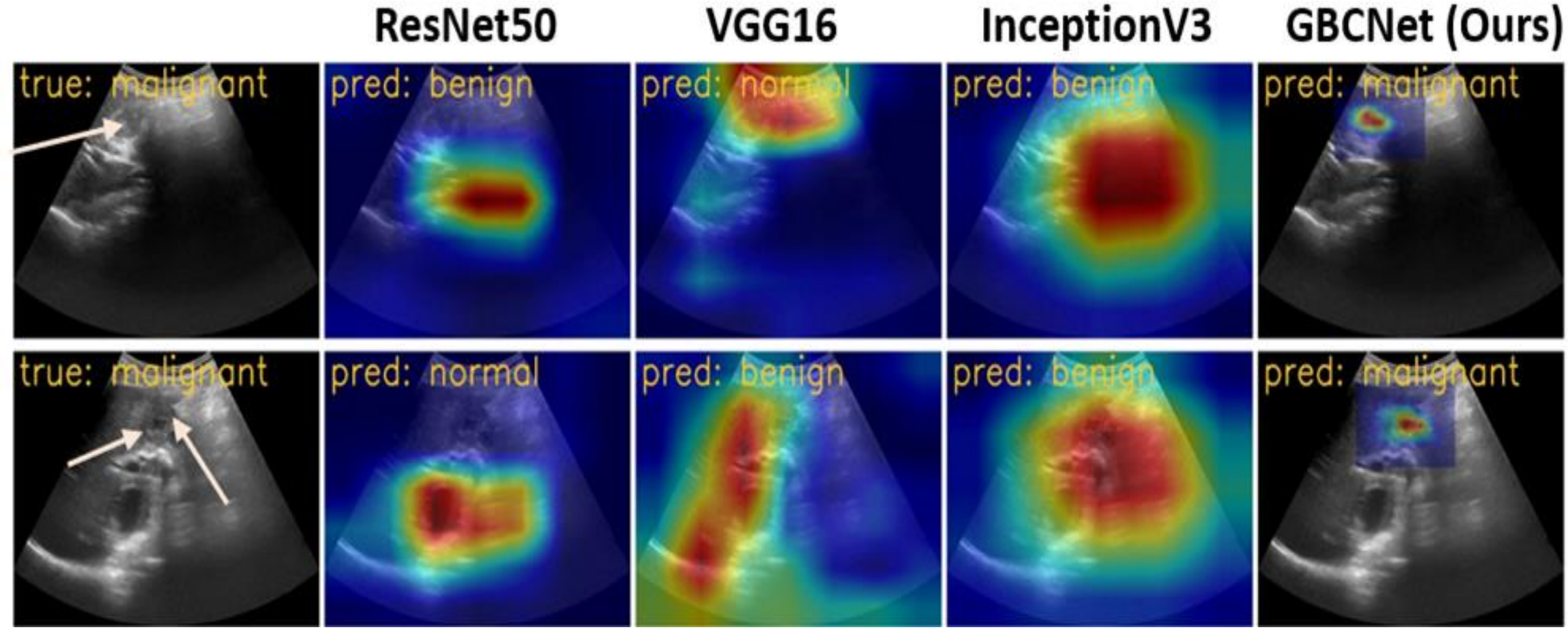
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Results

Method	Test Set				Cross Val.		
	Acc.	Acc.-2	Spec.	Sens.	Acc.	Spec.	Sens.
Radiologist A	70.0	81.6	87.3	70.7	–	–	–
Radiologist B	68.3	78.4	81.1	73.2	–	–	–
VGG16	62.3	72.1	90.0	38.1	69.3 ± 3.6	96.0 ± 4.6	49.5 ± 23.4
ResNet50	76.2	78.7	87.5	61.9	81.1 ± 3.1	92.6 ± 6.9	67.2 ± 14.7
InceptionV3	77.9	85.0	87.5	80.1	84.4 ± 3.9	95.3 ± 2.9	80.7 ± 9.7
Faster-RCNN	71.3	77.9	76.2	81.0	75.7 ± 5.3	84.0 ± 4.6	80.8 ± 10.4
RetinaNet	75.4	83.6	86.3	78.6	74.9 ± 7.3	86.7 ± 7.8	79.1 ± 8.9
EfficientDet	58.2	77.9	86.3	62.0	73.9 ± 8.4	88.1 ± 9.9	85.8 ± 6.1
GBCNet	87.7	91.0	90.0	92.9	88.2 ± 5.1	94.2 ± 3.7	$\mathbf{92.3} \pm \mathbf{7.1}$
GBCNet+VA	**91.0**	**95.9**	**95.0**	**97.6**	$\mathbf{92.1} \pm \mathbf{2.9}$	$\mathbf{96.7} \pm \mathbf{2.3}$	91.9 ± 6.3

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Qualitative Analysis

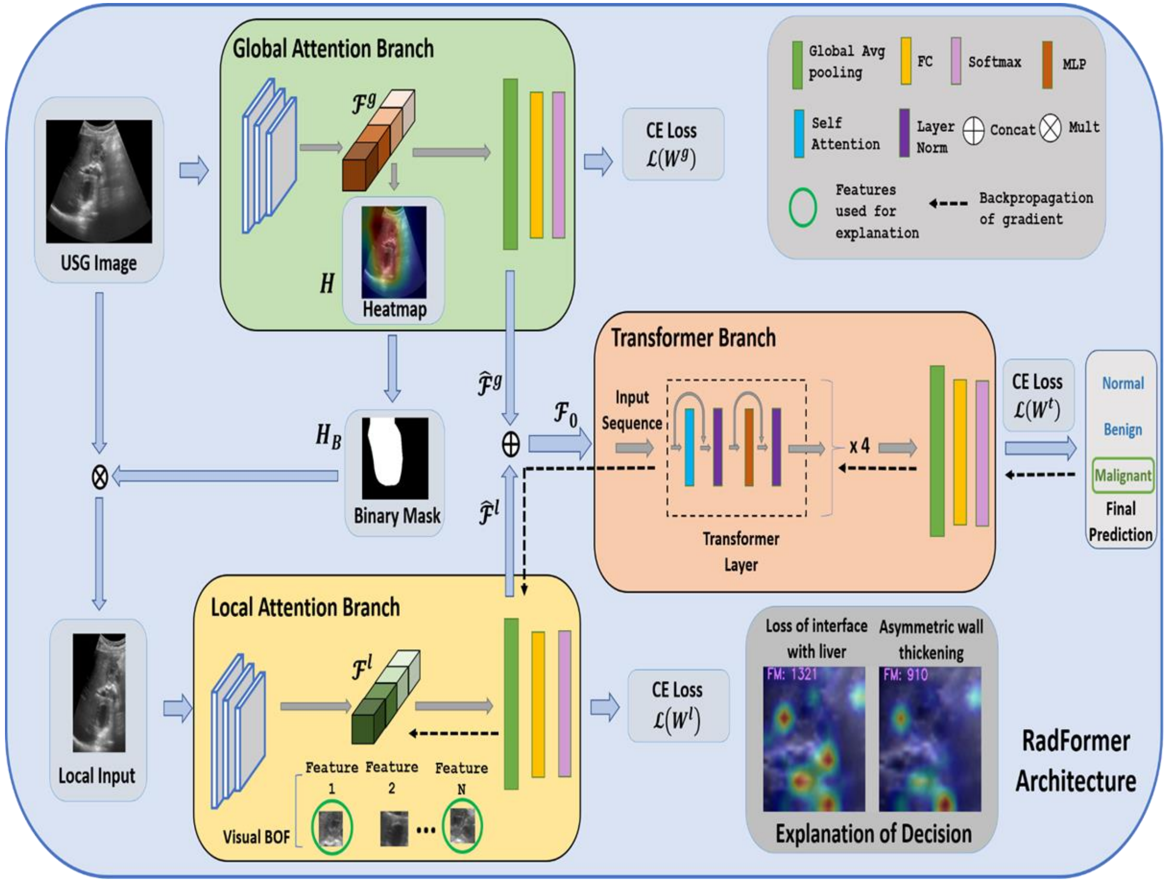


Interpretable Decision Making

(Elsevier Medical Image Analysis, 2023)

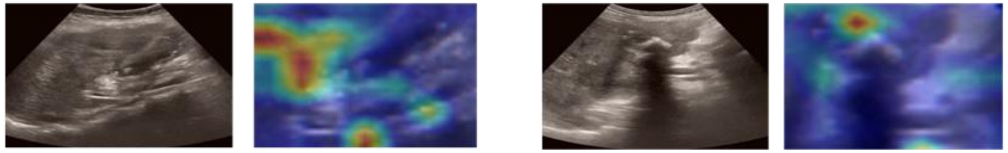
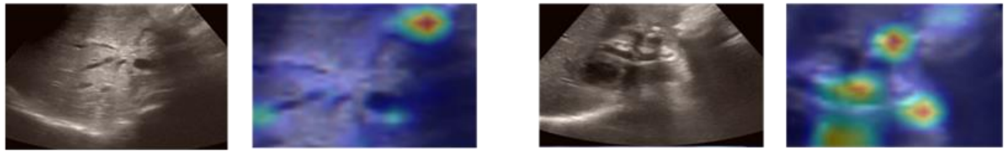
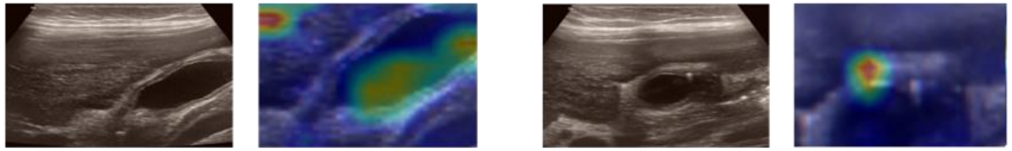
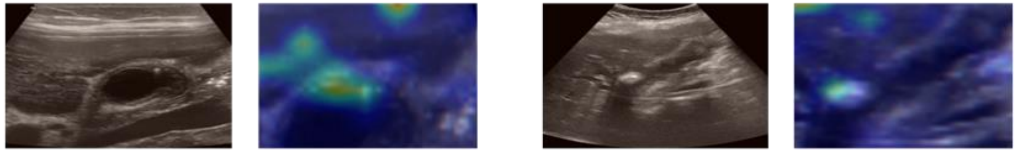
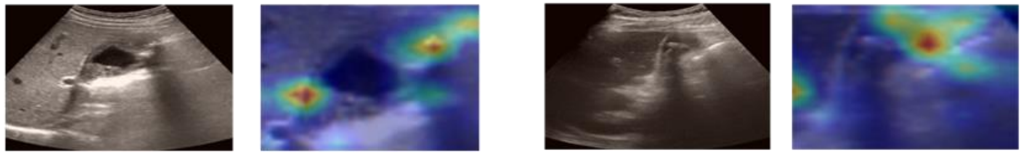
Interpretable GBC Detection

- Global-local attention based model
- Local branch - visual bag-of-feature style embedding
- Top visual words (gradient-weighted) are mapped with radiological features
- Explainable unit features consistent with RADS - compose radiologist-like explanations



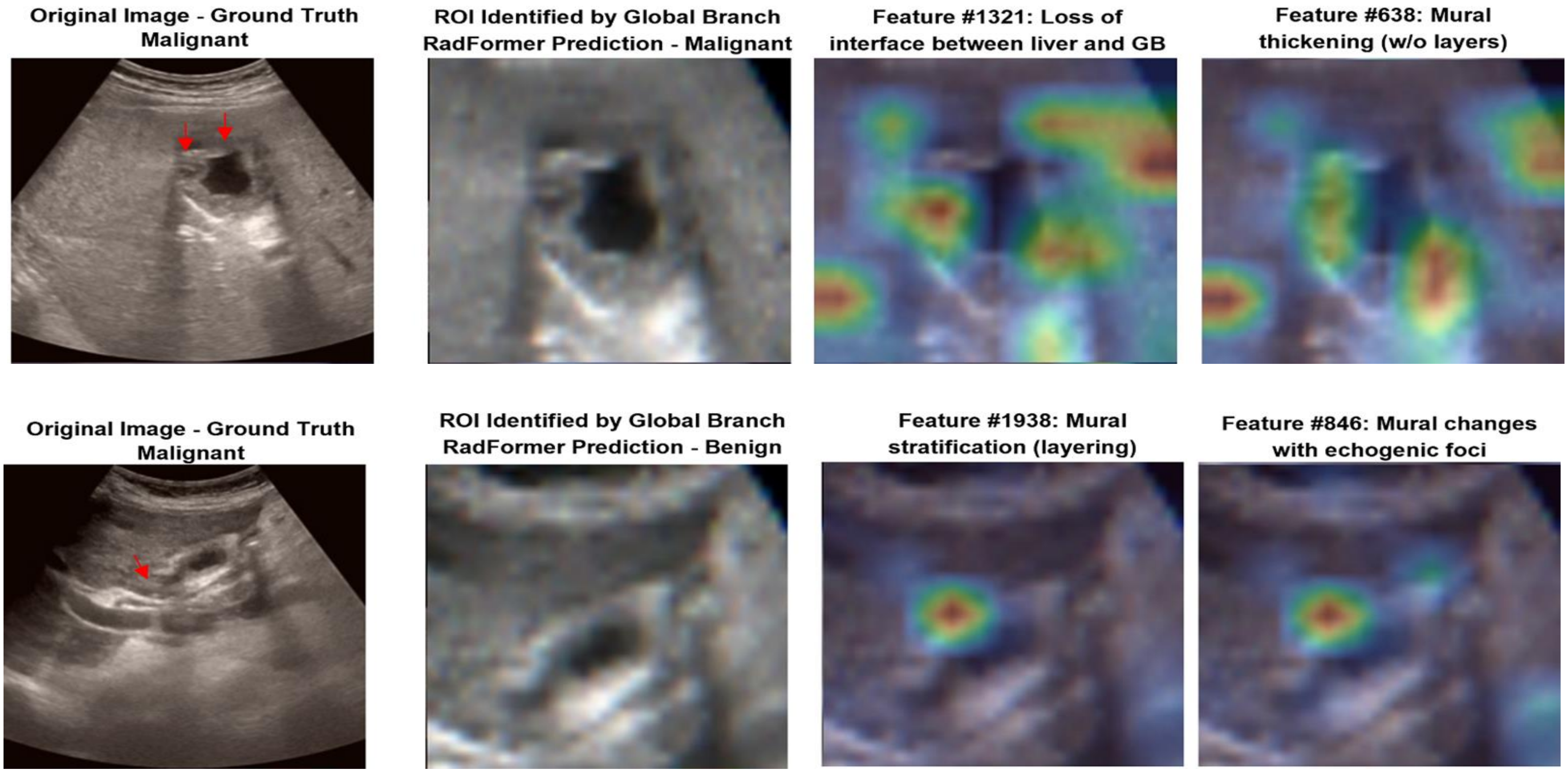
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Neural Features vs. RADS Lexicons

RadFormer feature id	GB-RADS lexicons	Sample images with activation visual of the feature from local BOF
1321	Loss of interface between GB wall and liver –significantly associated with malignancy	
1955, 1807, 1581, 1657	Extramural invasion – significantly associated with malignancy	
1935, 1359	Mural layering – intact inner and outer layers of the GB wall, favoring benign pathology	
879	Intramural echogenic mural foci – due to cholesterol deposition/ intramural calcification	
876	Intramural cyst within the GB wall – evidence of Rokitansky-Aschoff sinuses (RAS)	

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Sample Explanations



Learning from Unlabeled Videos

MICCAI 2022

Learning from Unlabeled Videos

- Labelled datasets are scarce for medical applications - specialized nature of annotations, data privacy

- Pre-training on natural image datasets - boosts performance, but not adequate due to domain gap

- Learn good representations for the downstream task from unlabeled video data

- Video data provides rich variation in viewpoint and natural temporal information - transformations of the same object across frames - learn effective representations.

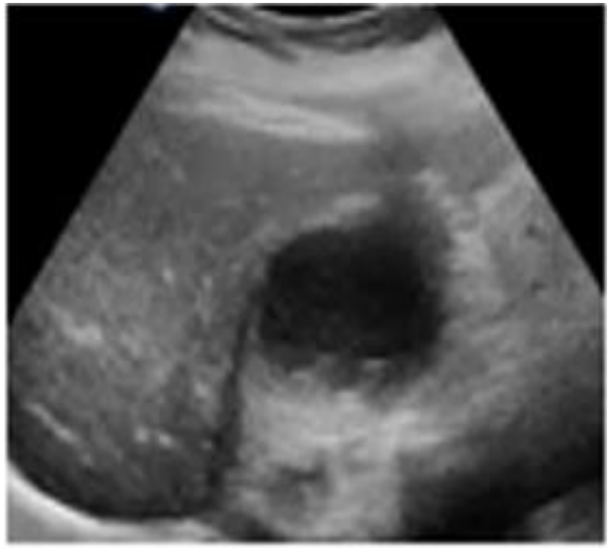
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Ideas and Contributions

- Previous SOTA techniques use only cross-video samples as negatives.
- USG is inherently different from natural videos
- Both inter-video and intra-video negatives are used - as opposed to SOTA - in a hardness sensitive curriculum
- Intra-video negatives are decided based on temporal distance
- Validated on two tasks - GBC, and COVID-19 detection.

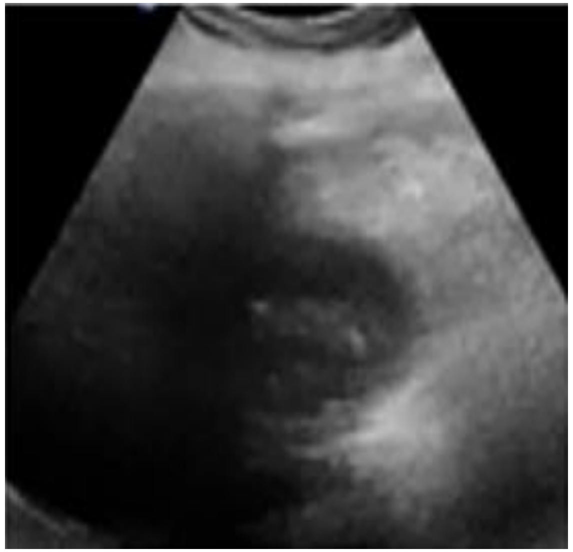
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Idea

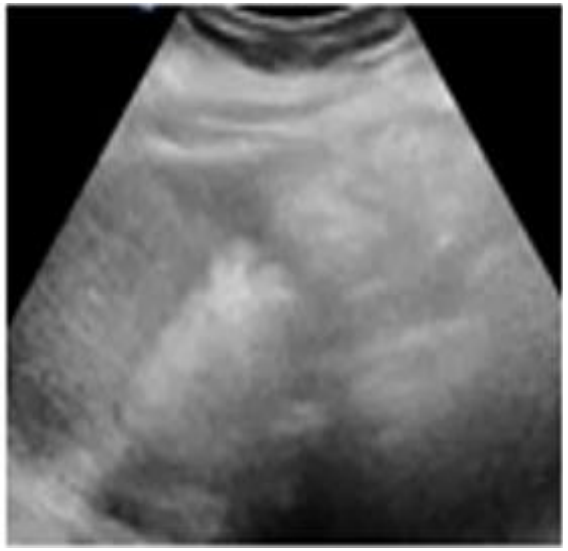


Malignant Sample
Stones and wall thickening

"Positive" Pair
Also shows wall thickening and stones



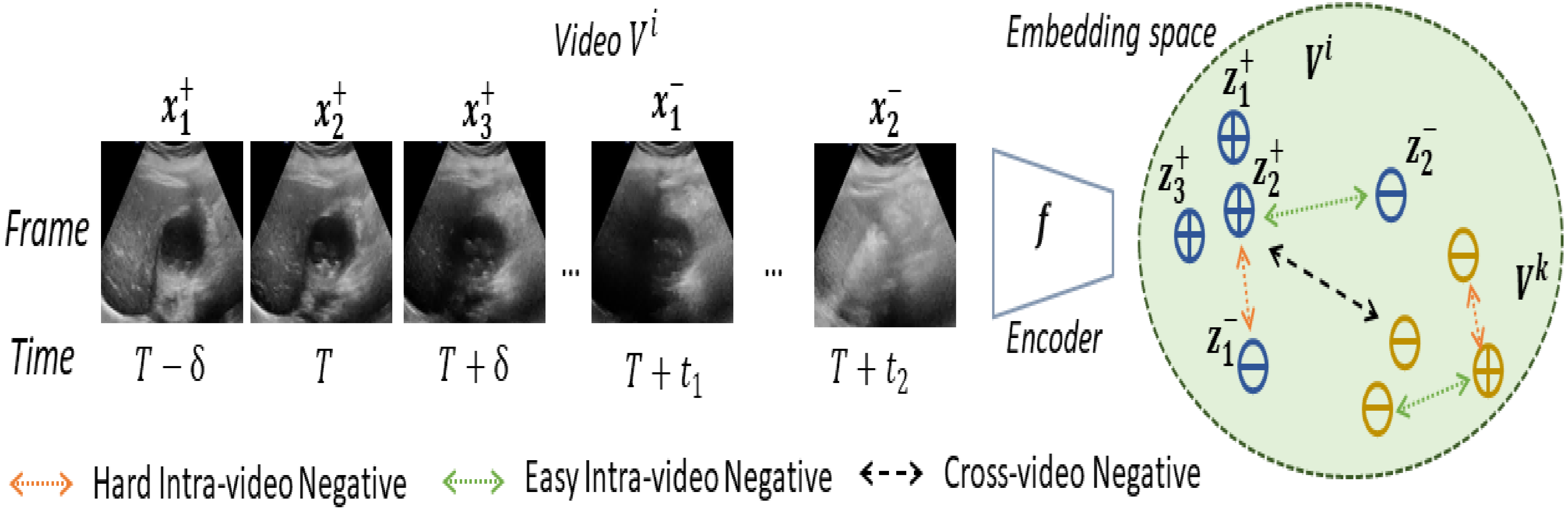
"Hard" Negative
Shows GB but no wall thickening



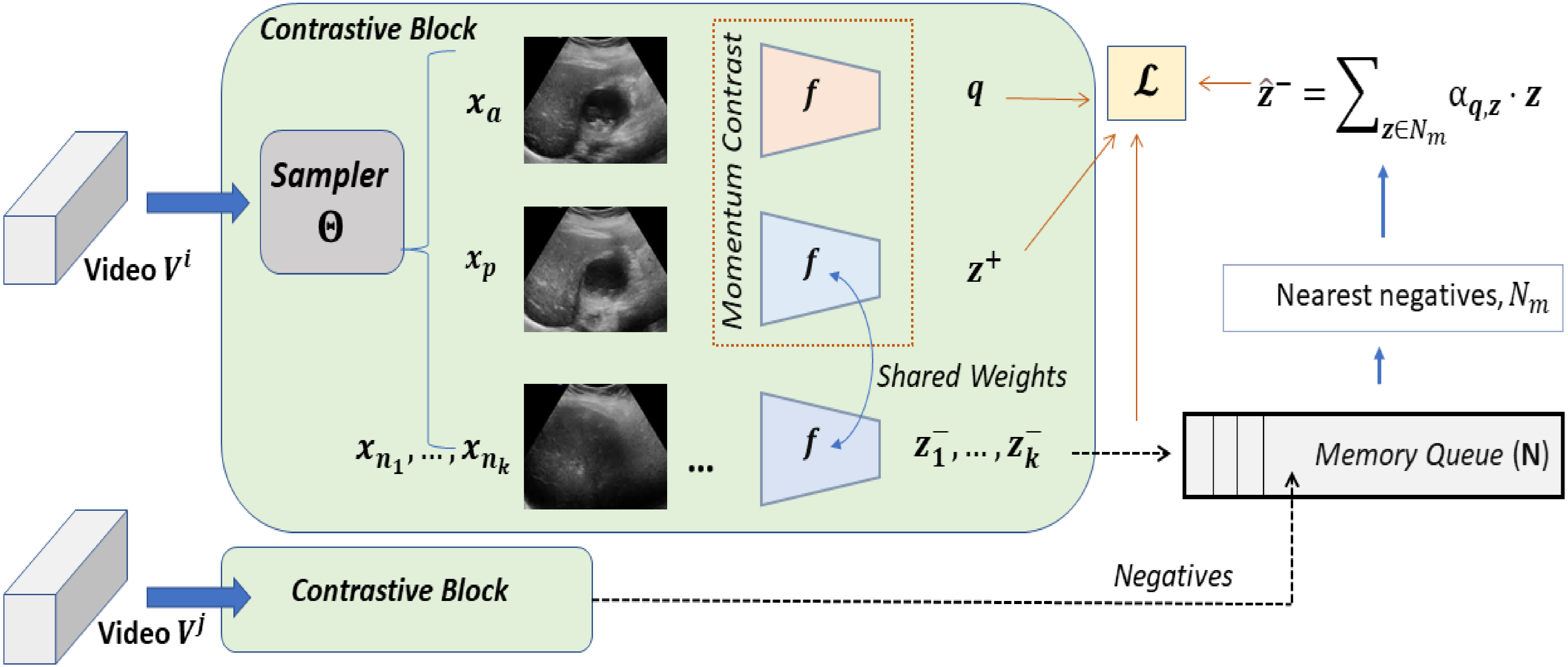
"Easy" Negative
Even the GB is not visible

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Idea



Proposed Pre-training Overview



Downstream Task

- Gallbladder Cancer detection from abdominal US images
- Pretrain on Unlabeled GB videos
  - 64 videos, 15800 frames (also our contribution)
- Finetune on GBCU dataset (GBCNet: CVPR 22)
  - Non-malignant: 990, Malignant: 265
- Proposed "Easy-Hard" curriculum also validated on public COVID 19 dataset.

Results

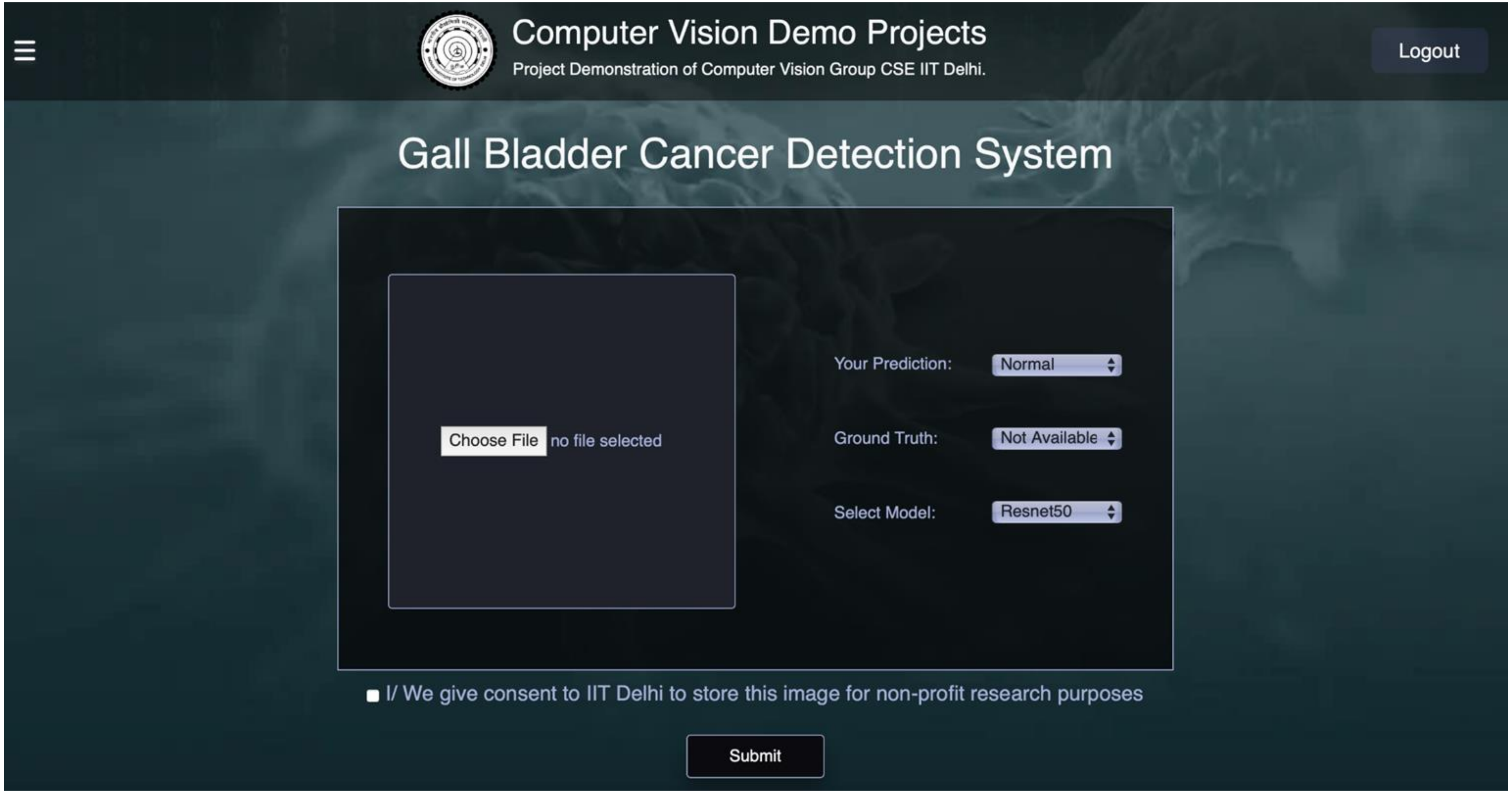
- Results for ResNet50 classifier model with different pre-training

Method	Acc.	Spec.	Sens.
Pretrained on [14]	0.867 ± 0.070	0.926 ± 0.069	0.672 ± 0.147
SimCLR [10]	0.897 ± 0.040	0.912 ± 0.055	0.874 ± 0.067
SimSiam [11]	0.900 ± 0.052	0.913 ± 0.059	0.861 ± 0.061
BYOL [15]	0.844 ± 0.129	0.871 ± 0.144	0.739 ± 0.178
MoCo v2[19]	0.886 ± 0.061	0.893 ± 0.078	0.871 ± 0.094
Cycle-Contrast [25]	0.861 ± 0.087	0.867 ± 0.098	0.844 ± 0.097
USCL [12]	0.901 ± 0.047	0.923 ± 0.041	0.831 ± 0.072
Ours	$\mathbf{0.921 \pm 0.034}$	$\mathbf{0.926 \pm 0.043}$	$\mathbf{0.900 \pm 0.046}$

Current/Future Work

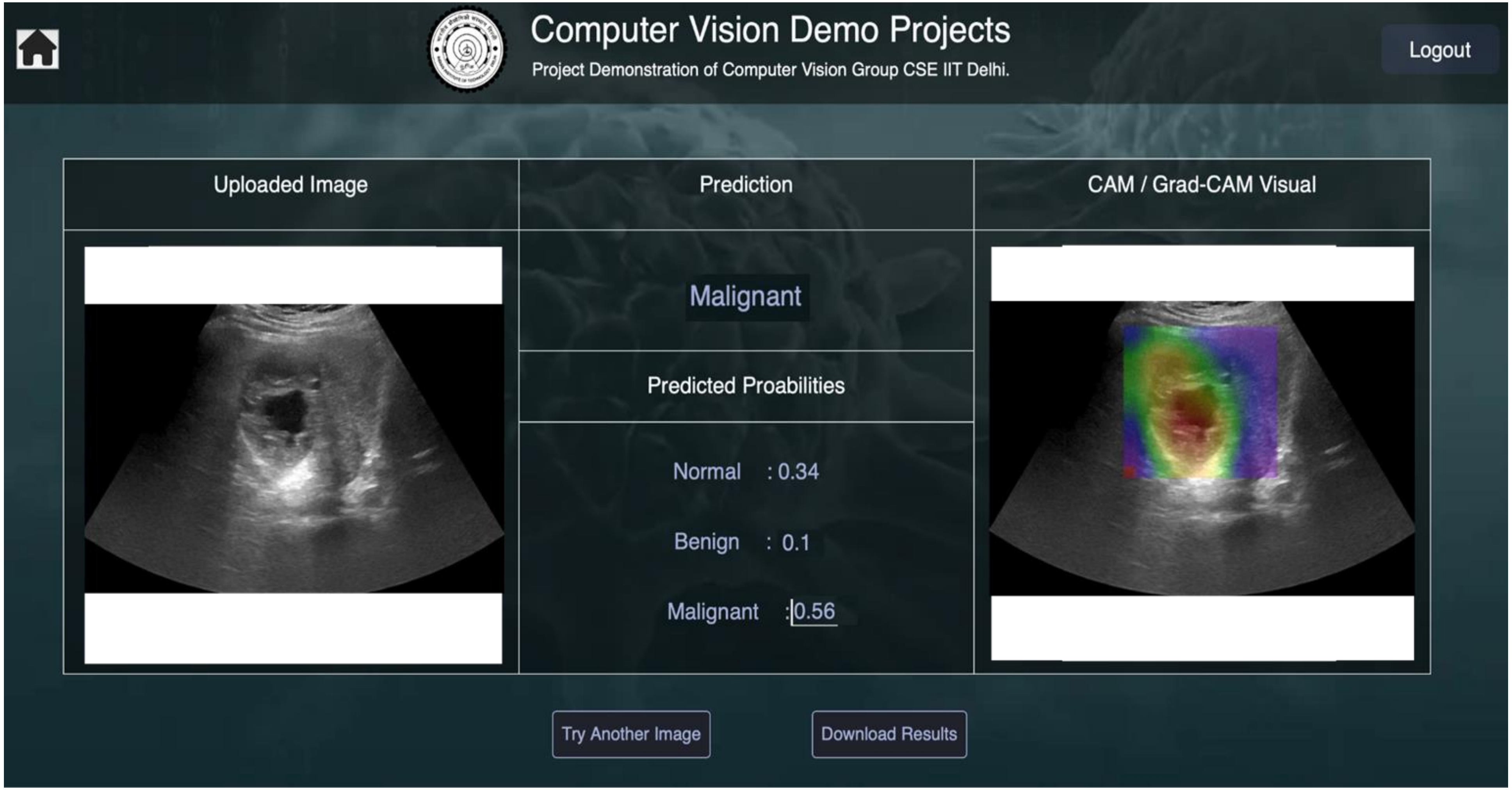
- WebApp/ IoT Device for onsite (hospital) detection: transformational potential
- Video-based Detection – localize anomaly
- Generalized models: tackle domain shift for different hospitals
- CT-based detection
- Her2Neu detection from CT for targeted therapy

Web-demo


Computer Vision Demo Projects
Project Demonstration of Computer Vision Group CSE IIT Delhi.
Logout
Gall Bladder Cancer Detection System
Choose File no file selected
Your Prediction:
Normal
Ground Truth:
Not Available
Select Model:
Resnet50
I/ We give consent to IIT Delhi to store this image for non-profit research purposes
Submit

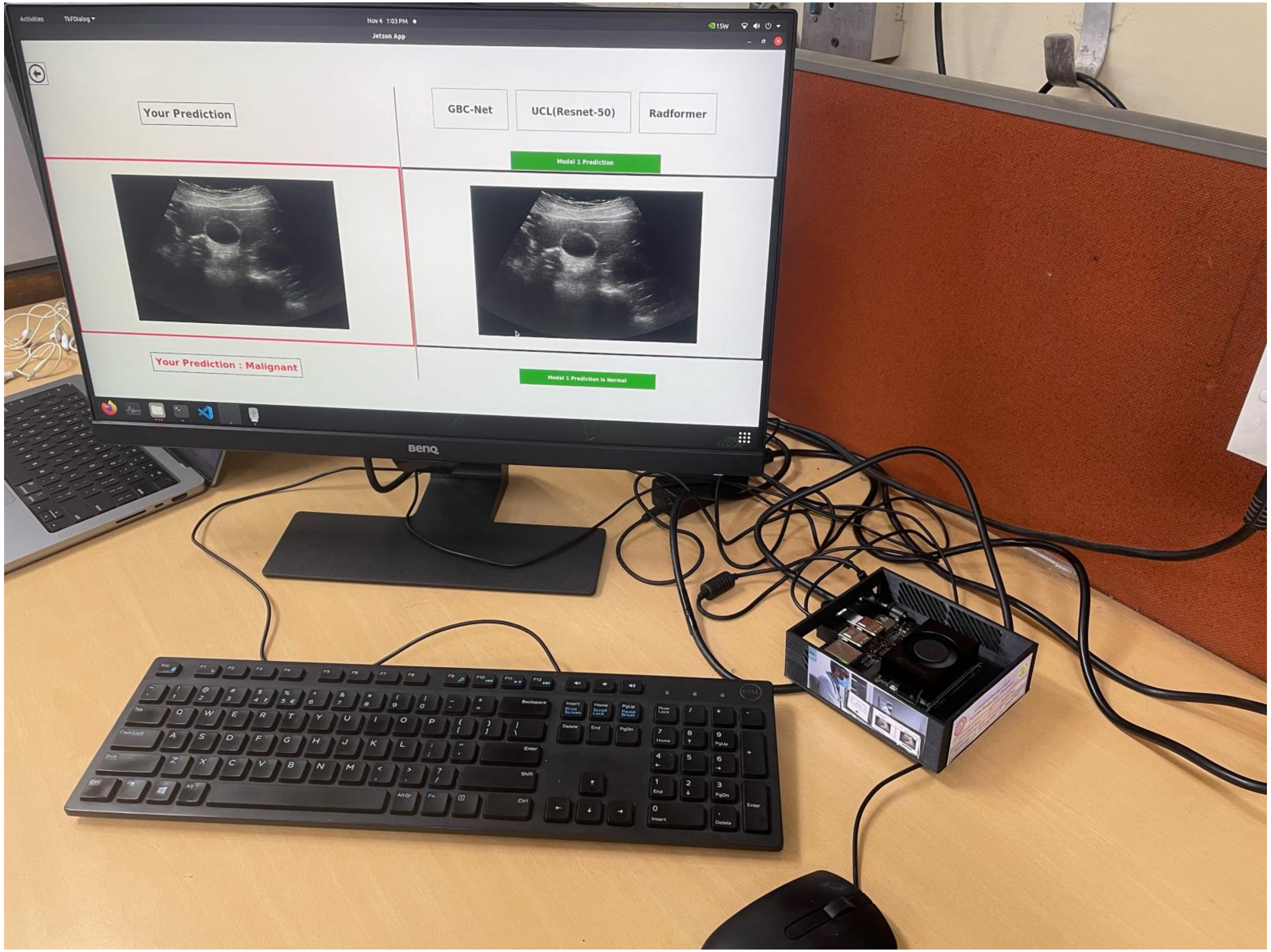
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Web-demo



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

POC Device



Collaborators

Soumen Basu
PhD, IITD

Mayank Gupta
MSR, IITD

Dr. Usha Dutta
PGIMER.

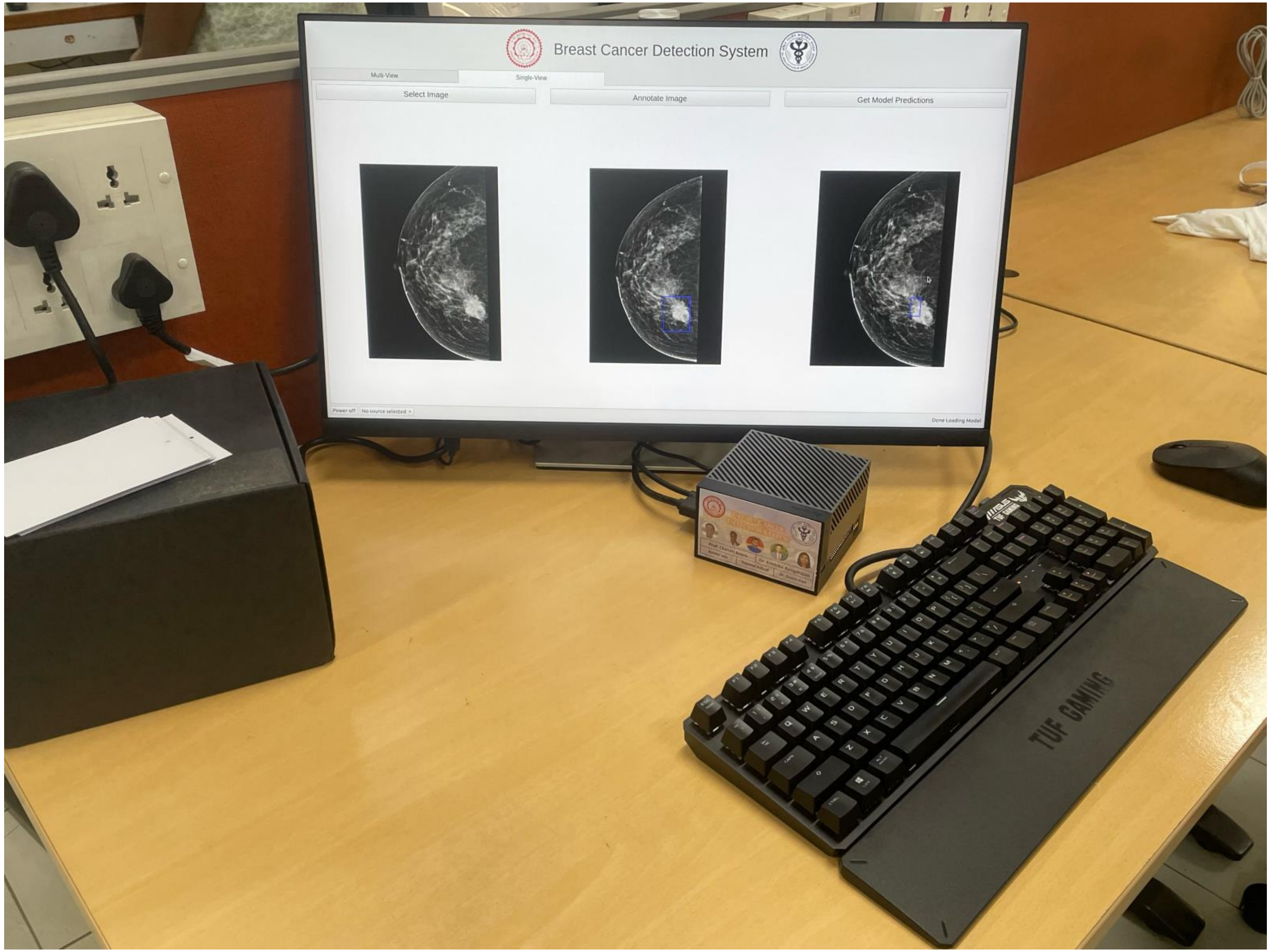
Dr. Pankaj Gupta
PGIMER
Also: PhD, IITD

And many more from PGIMER Chandigarh, and IIT Delhi…

Breast Cancer Detection from Mammograms

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

POC Device



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Motivation

- On an average, cancer is detected at stage 1 or stage 2 in the West, however in India it is detected in stage 3 or stage 4

- However, mammography is expensive, it is a dedicated machine only for breast imaging, requires significant expertise to perform and interpret, requires highly trained breast radiologists

- Implementation of population wide screening is impossible in India, given the resource constraints and the dismal doctor-patient ratio (1:834). Radiologist-patient ratio is even more skewed at (1:1,00,000)[1].

1. Arjun Kalyanpur, Commentary - Radiology in India: The Next Decade. Indian J Radiol Imaging. 2008

Motivation

- Even those who do become radiologists, can do so without ever seeing a single mammogram

- Thus, breast cancer in India
  - Has unique challenges (Earlier onset, possibly different biology)
  - Common disease with high mortality
  - Presents an opportunity to reduce mortality, proven screening tool, unlike most other cancers

- We asked if AI could bridge this gap

Our work so far

- To train a neural network to achieve high specificity and negative predictive value for cancer detection on mammograms
  - Small Mass Detection
  - Isodense Mass Detection
  - Multimodal Detection
  - Multiview Detection
- To develop and test a tool for simulation training in mammography
- To develop and test a tool for rapid report generation, which can simultaneously generate annotated data for training deep learning models.

Collaborators

Kshitiz Jain
MSR, IITD

Tajamul Ashraf
MSR, IITD

Dr. Krithika Rangarajan
AIIMS Delhi

Prof. S. Banerjee
Ashoka

And many others from AIIMS Delhi and IIT Delhi…

Thank you!

Extra Slides

Small Mass Detection in Breast Cancer

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Small Mass Detection

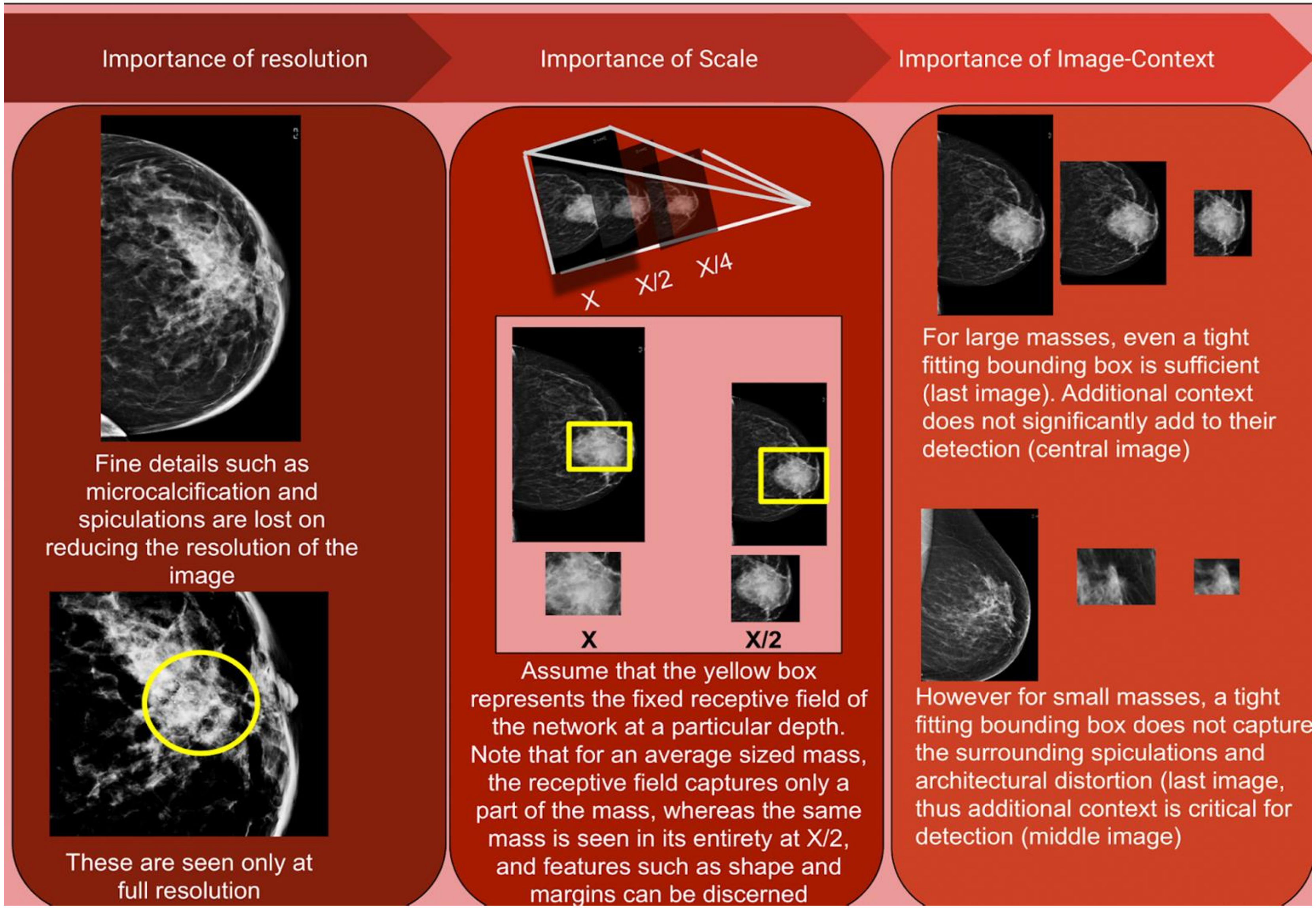
- It is essential to detect cancers while they are still small. 10-year Survival of patients fall from 95 percent to 60 percent when cancer size increases from 1 cm to 3 cms, some cancers are very fine microcalcifications

- This is the reason why mammography has the highest spatial resolution of all radiological modalities: 2300X1800 to 4096X 5625 pixels.

- Training a neural network at such high resolution can be difficult due to memory constraints. Most models reported in literature used input size of 1024X1024 pixels, leading to significant loss of resolution

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

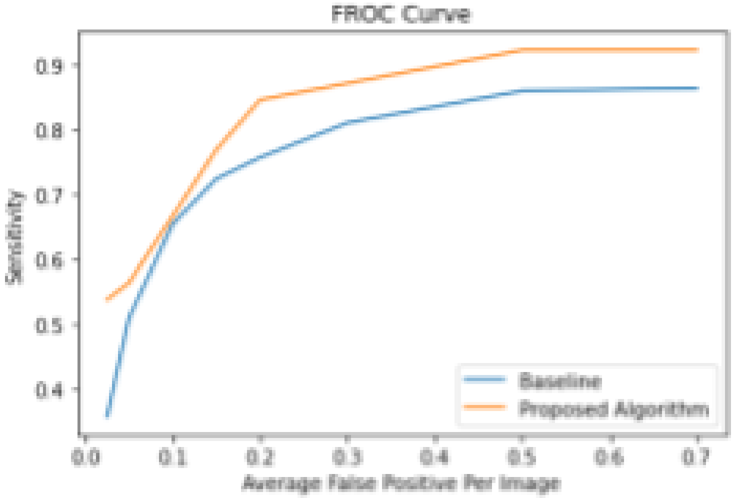
Small Mass Detection



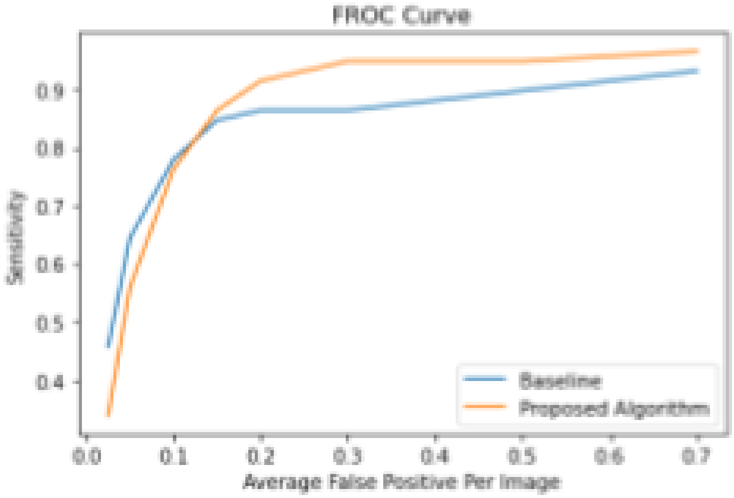
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Small Mass Detection

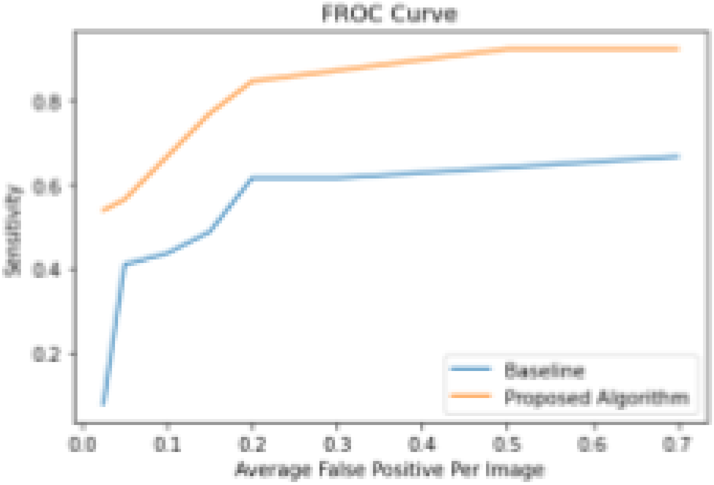
Dataset	**Sensitivity at 0.15 FPI (proposed/ baseline)**	**0.2 FPI (proposed/ baseline)**	**0.3 FPI (proposed/ baseline)**
Diagnostic mammography	0.7037/*0.6543*	0.7818/*0.6831*	0.8353/*0.7201*
Screening Mammography (External Dataset)	0.8644/0.8474	0.9152/0.8644	0.9491/0.8644
Small Mass dataset	0.7692/ 0.4870	0.8461/0.6153	0.8717/0.6153



a)DM dataset

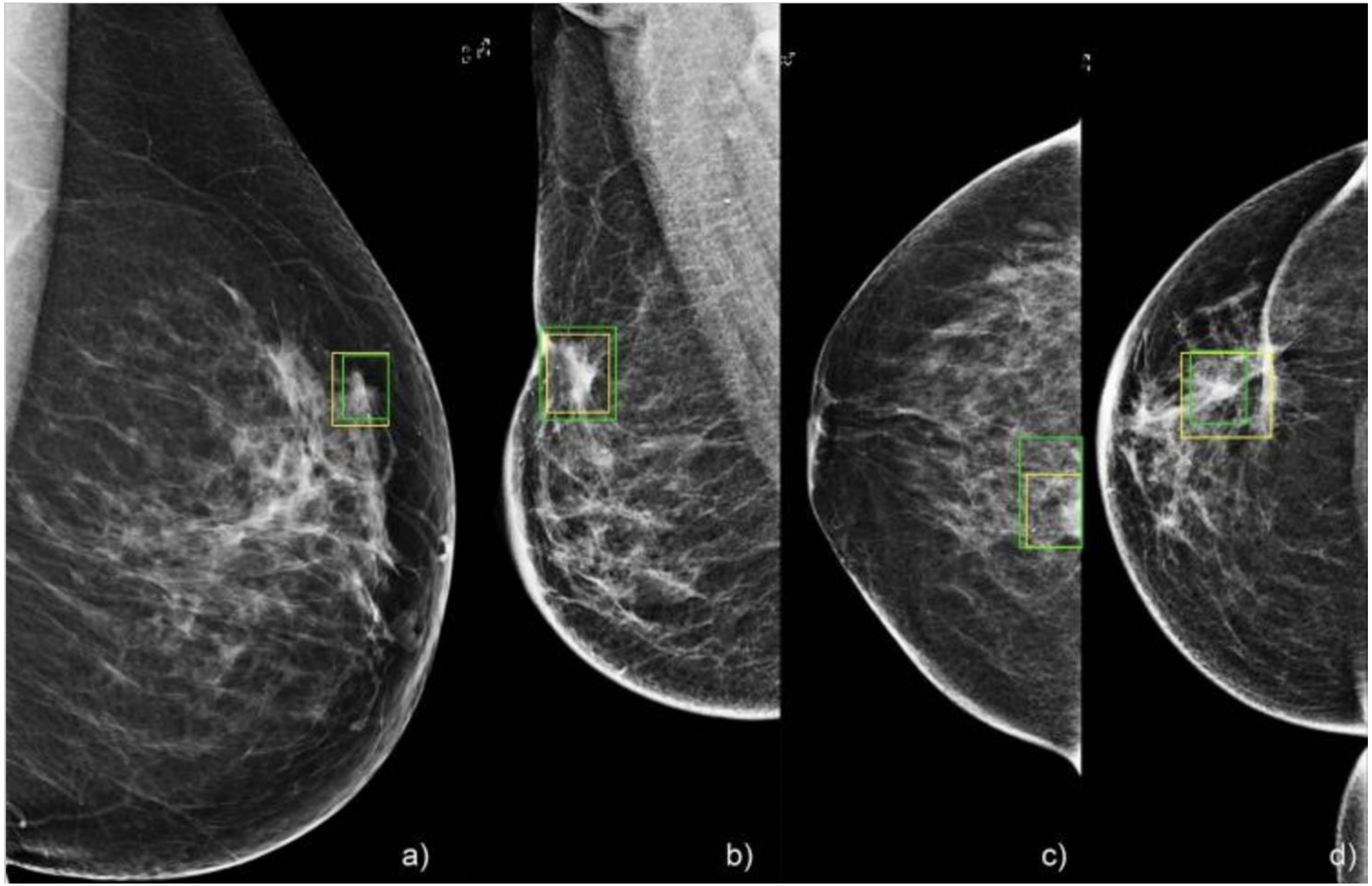


b) SM dataset



c)Small Cancer dataset

Small Mass Detection (Green: GT, Yellow: Pred.)



Breast Cancer Detection in Dense Breasts

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Breast Cancer Detection in Dense Breasts

- Mammographically dense breasts have significantly higher risk of cancer, while also having lower accuracy for cancer detection by mammography
- Lower patient age is associated with denser breasts. Given the lower age of cancer in India, it is essential the network works well on dense breasts
- We found that our models had lower accuracies in detection of isodense obscure masses in dense breasts.

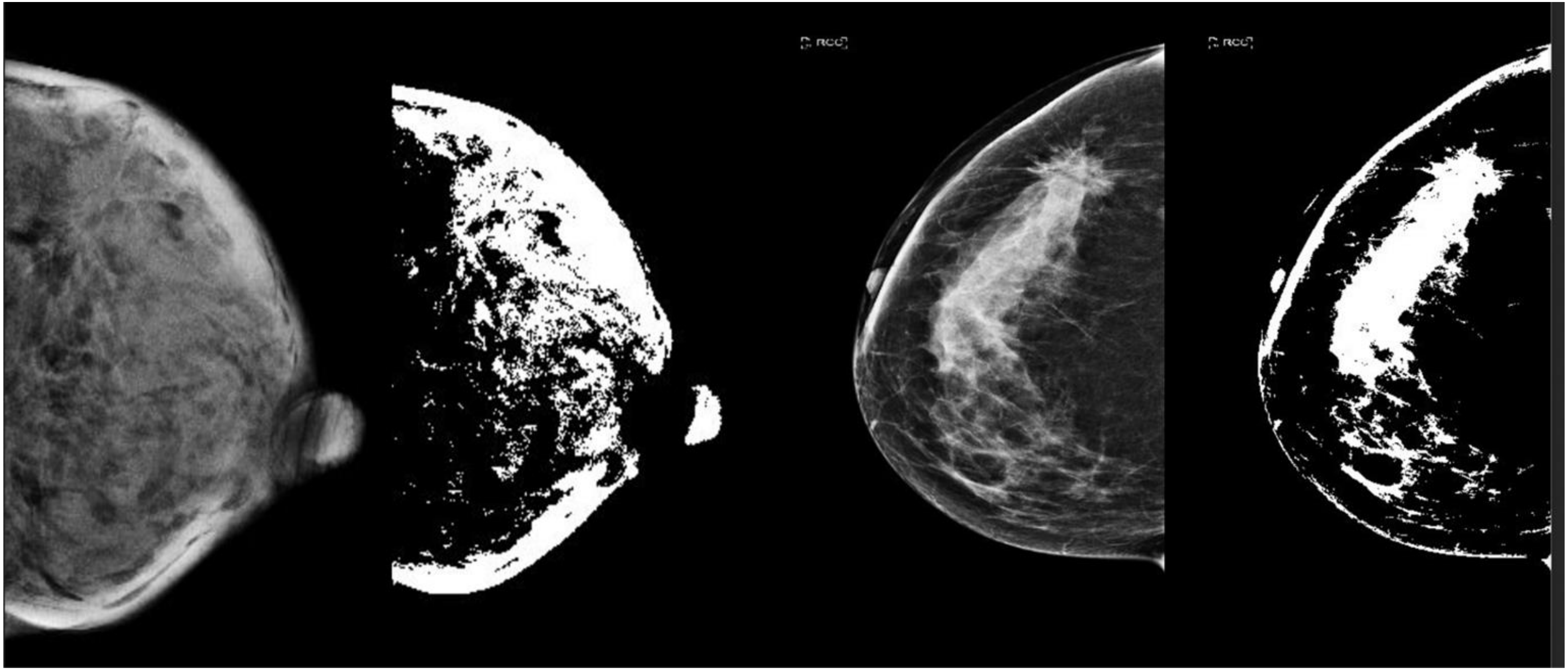
Key Idea

Followed the approach of a radiologist

- Typically, radiologists focus on architectural changes in the breast
- They adjust the contrast on the screen to look at the mass better
- They look at the opposite breast to look for regions of asymmetry

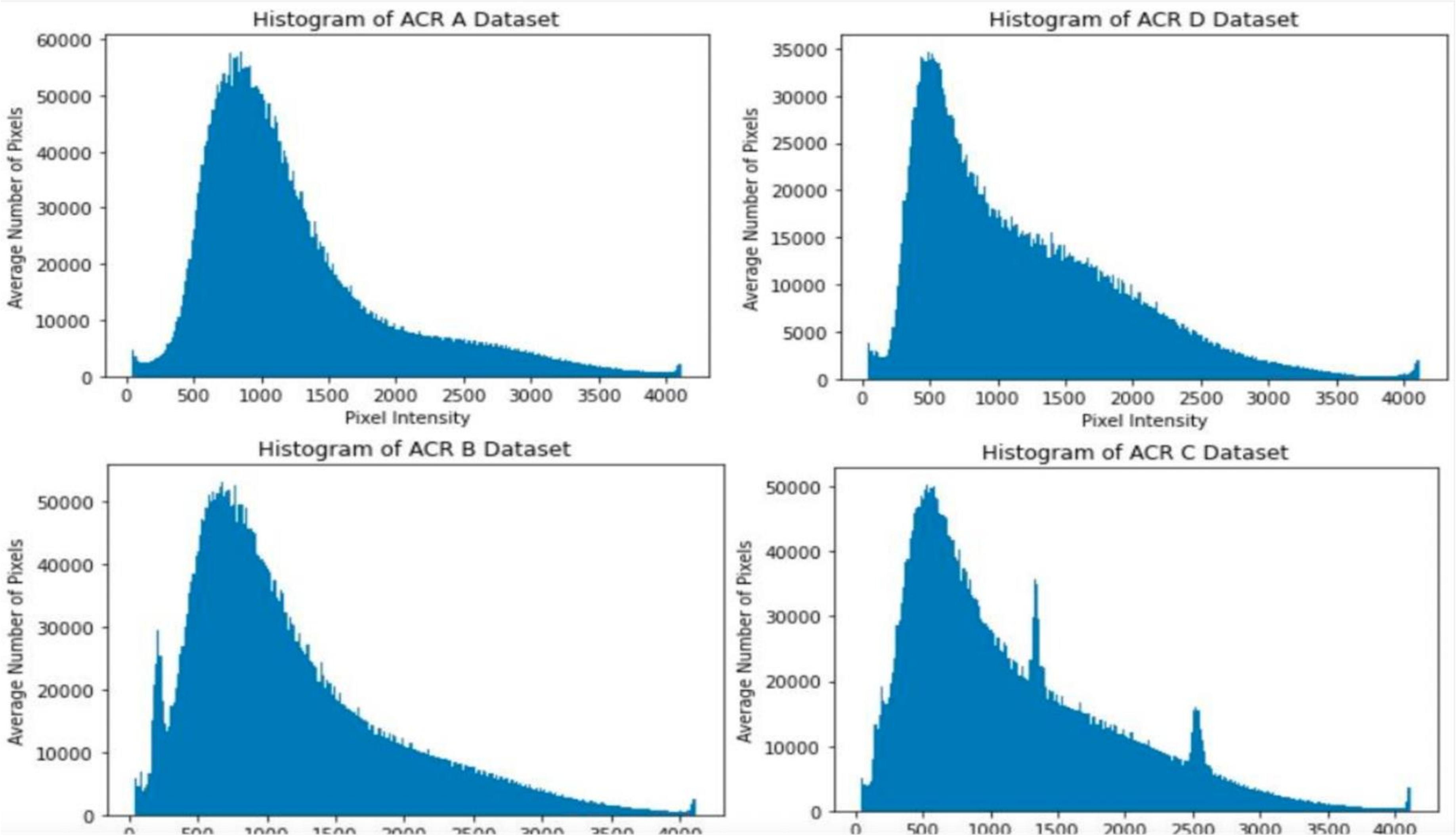
We took these factors into design of model

Key Idea: Focus on Architectural Changes



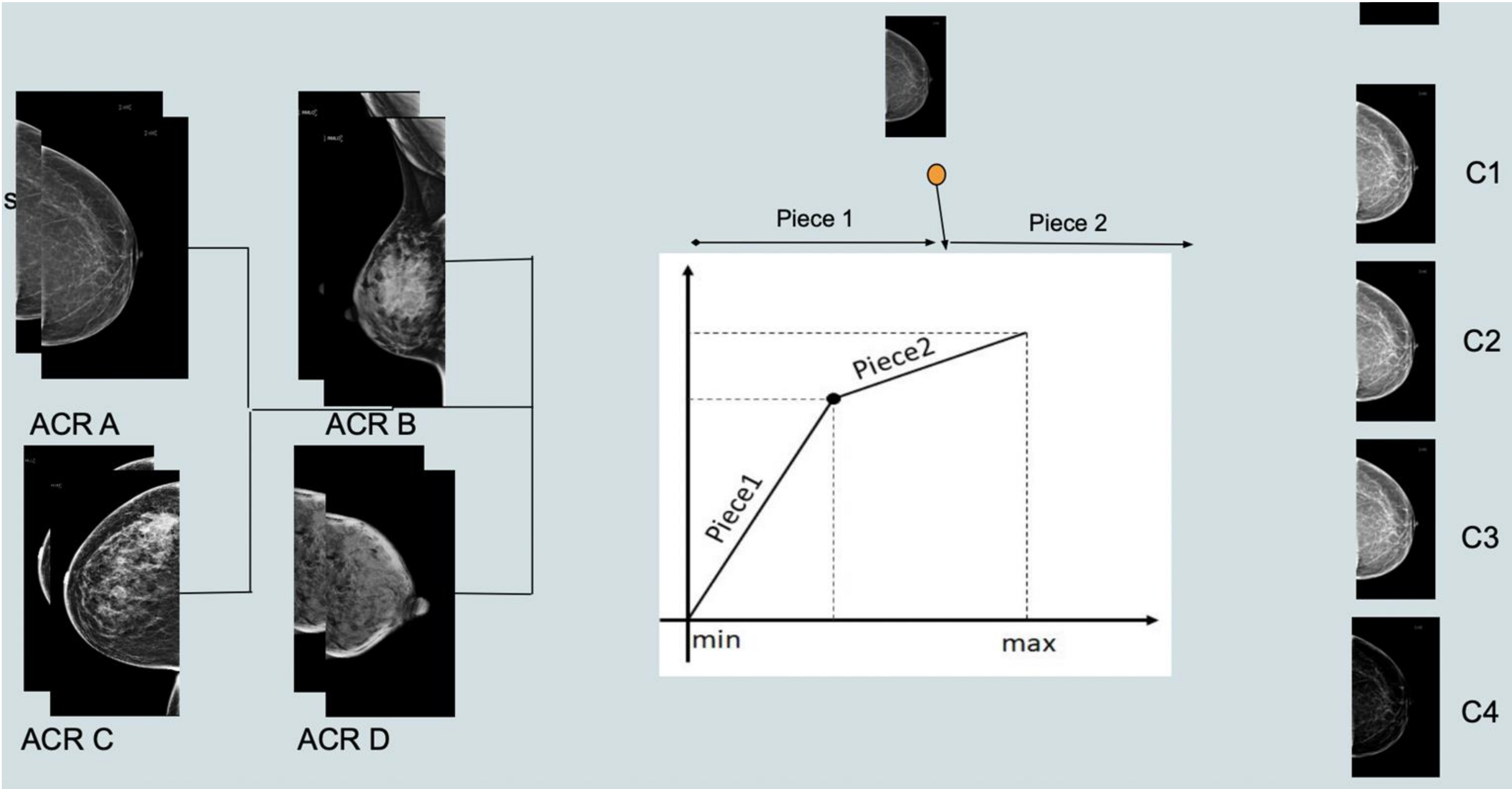
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Idea: Present at Different Contrasts



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Key Idea: Present at Different Contrasts



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

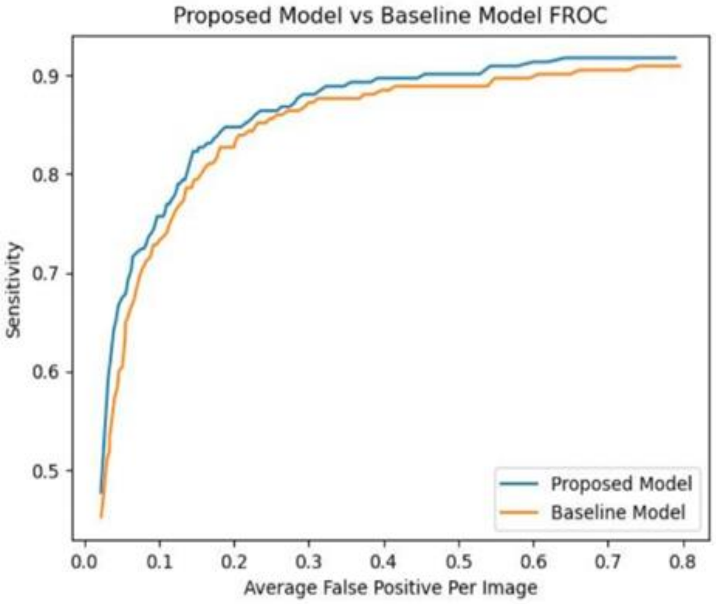
Results

Table3

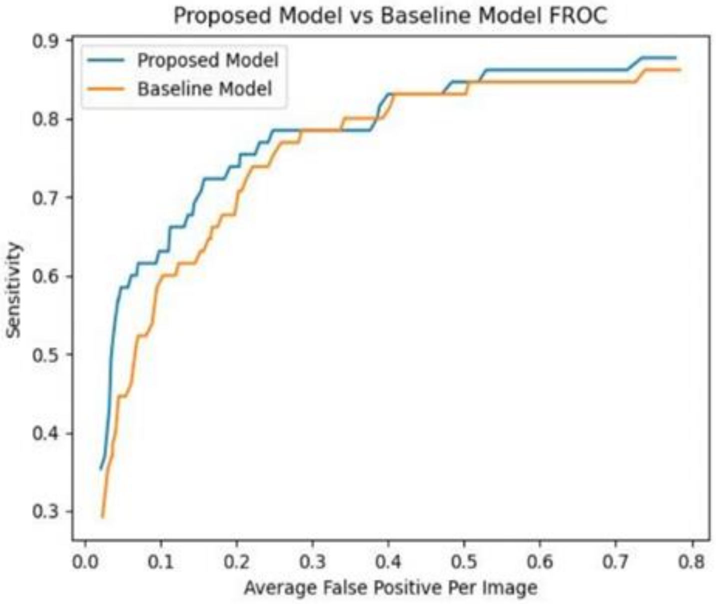
FPI	0.025	0.05	0.1	0.15	0.2
Baseline	0.373	0.386	0.626	0.653	0.746
TI	0.373	0.386	0.626	0.653	0.746
CABD	0.453	0.560	0.626	0.786	0.840
TI+CABD	0.546	0.613	0.760	0.773	0.813
Bilateral	0.240	0.413	0.733	0.800	0.826
Proposed	0.453	0.560	0.666	0.840	0.853

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

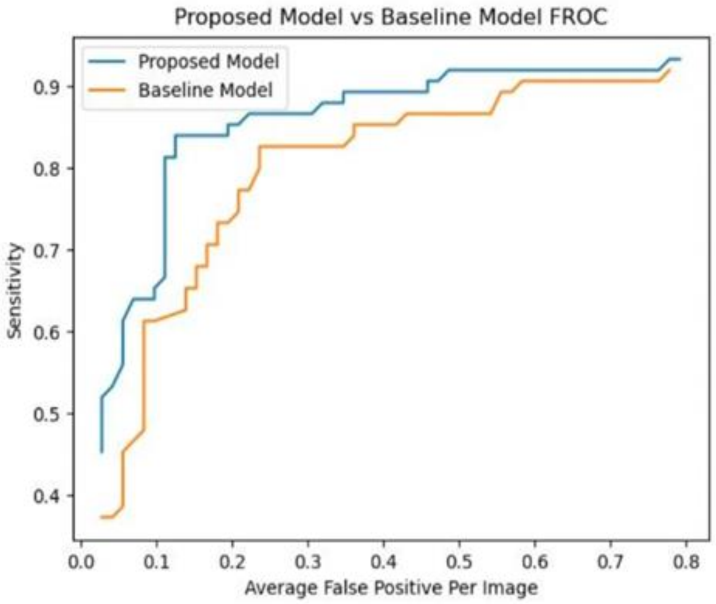
Results



a) Full DM dataset



b) Dense breast subset DM dataset



c) Isodense/Obscure mass subset breast subset DM dataset

Results

Authors	Sensitivity @ FPI
Kozegar et al (11)	0.87 @ 3.67
Akselrod-Ballin et al (12)	0.93 @ 0.56
Dhungel et al (13)	0.95 @ 5
Ribli (14)	0.79@0.1 & 0.90 @ 0.3
Richa Agarwal et al (15)	0.9 @ 0.44
Proposed	0.86 @ 0.1 (+/- 0.069) 0.90 @ 0.2 (+/- 0.065) 0.93 @ 0.44 (+/-0.022)

Subgroup Analysis for GBC

Lancet Regional Health - SE Asia (Accepted)

Our DL Model – Prospective Diagnostic Study

- Large prospective study – Used our MS-SoP model for automatic detection of GBC at abdominal US

- Compared diagnostic performance with two expert radiologists.

- Performed subgroup analysis to demonstrate the robustness of DL-model.
  - Polyps and mural thickening,
  - Contracted gallbladder
  - Neck lesions, etc.

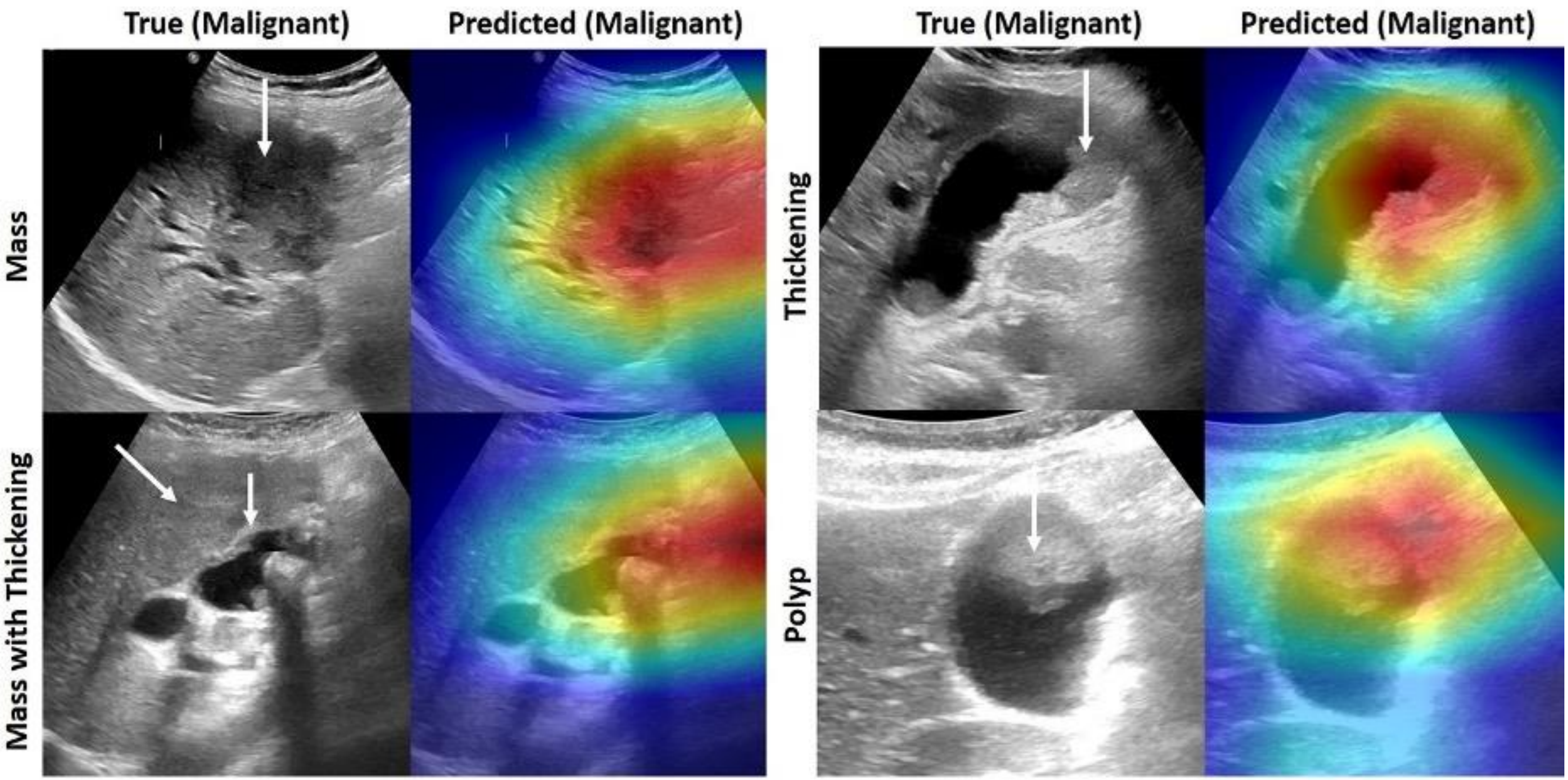
- 565 prospective patients at PGIMER: Train: 233, Val: 59, Held out Test: 273

Data Acquisition

- Radiologists with 1-8 years post-training experience in the abdominal US performed GB US on Logiq S8 scanner
- Convex transducer with a frequency of 1-5 MHz after at least 6 hours of fasting.
- Independent reading by 2 radiologists with 2 years and 8 years of post-training experience in the abdominal US.
- The radiologists were aware that the patients had GB diseases but were blinded to the findings of the previous imaging tests and the final diagnosis.

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

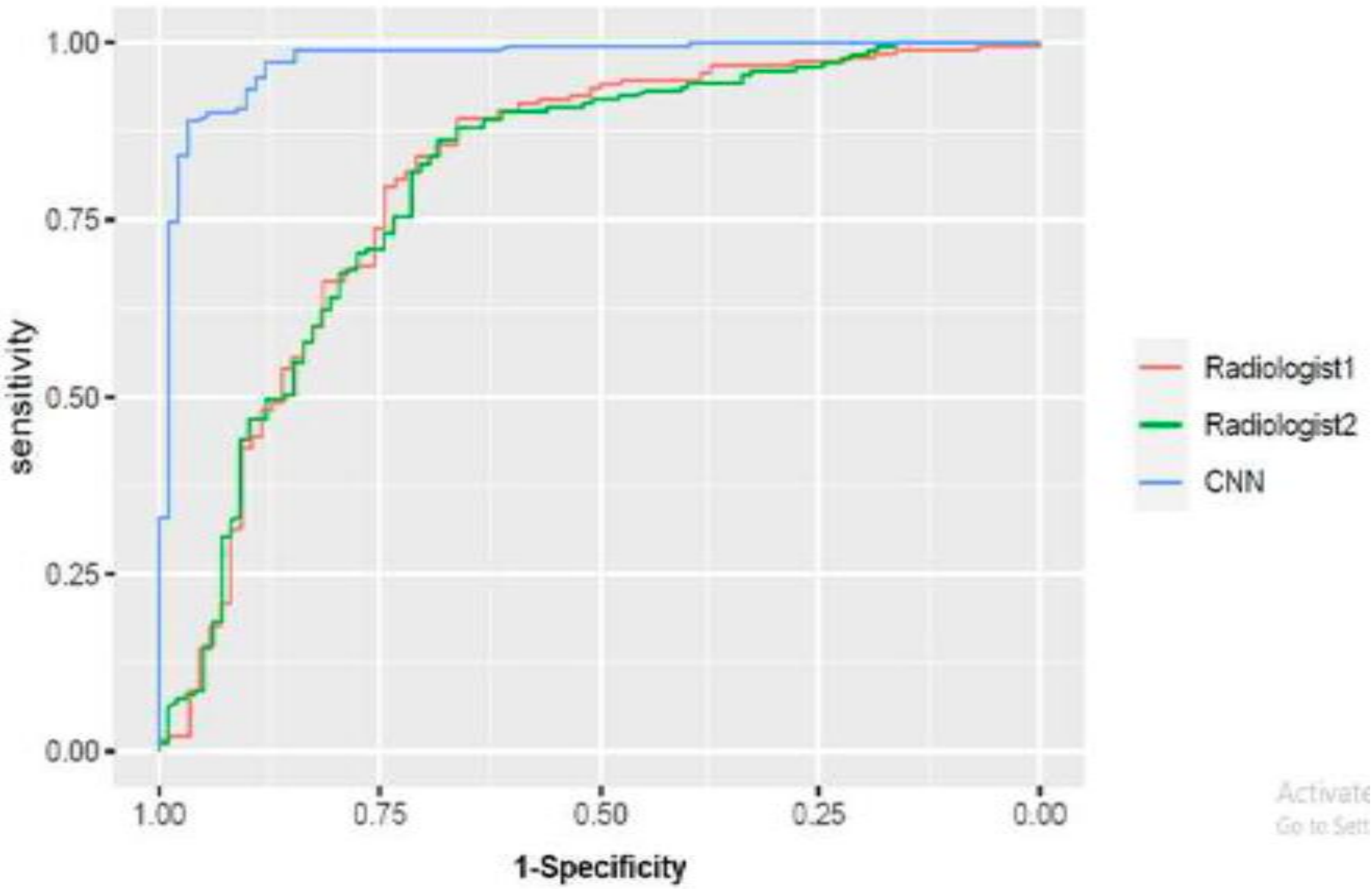
Output: Different Morphological Subtypes



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

AUC: DNN Vs Radiologist

- 1- Specificity = Probability that a true negative will test positive.
- 1- Specificity = FP / N
- Also referred to as False Positive Rate (FPR) or False Positive Fraction (FPF).



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Performance in Various Subgroups

Groups	% Sensitivity (95%CI)	% Specificity (95%CI)	% PPV (95%CI)	% NPV (95%CI)	% Accuracy (95%CI)	AUC (95%CI)
			Overall			
CNN	92.3 (88.1-95.6)	74.4 (65.3-79.9)	90.1 (84.9-94.1)	80 (70.2-87.6)	86.4 (82.2-90.5)	0.887 (0.844-0.930)
Radiologist 1	86.8 (81.1-91.4)	67 (56.3-76.5)	87 (81.31-91.5)	76.1 (65.8-84.5)	80.2 (75-84.8)	0.826 (0.767-0.884)
Radiologist 2	87.9 (82.3- 92.3)	80 (70.2- 87.7)	89.7 (84.32-93.8)	75.2 (65.4-83.4)	85.3 (80.5- 89.3)	0.837 (0.781-0.892)
			Stones			
CNN	92.2 (87-95.2)	79.6 (71.9-93.1)	90.1 (82.5-95.1)	80.0 (67.0-89.5)	87.8 (82.3-93)	0.890 (0.836-0.945)
Radiologist 1	90.2 (82.7-95.2)	72.2 (58.4- 83.5)	85.5(77.3-91.7)	76.9 (63.1-87.4)	83.9 (77.3- 89.4)	0.812 (0.733-0.891)
Radiologist 2	90.1 (82.5- 95.2)	77.8 (64.4-87.9)	88.24(80.3-93.7)	81.1 (68-90.5)	85.8 (79.3-90.9)	0.835 (0.761-0.909)
			Mass			
CNN	98.2 (90.4-99.9)	100 (2.5-100%)	99.1 (95.1-99.9)	20 (0.5-71.6)	98.2 (90.6-99.6)	1
Radiologist 1	96.4 (87.6- 99.5)	100 (2.5-100)	100 (93.4-100)	25 (0.6-80.6)	96.5 (87.9-99.6)	1
Radiologist 2	100 (93.6- 100)	100 (2.5-100)	100 (93.6-100)	100 (2.5-100)	100 (93.7- 100)	1
			Thickening			
CNN	87.8 (78.7-93.9)	74.1 (64.4-84.2)	84.1 (74.7-91)	86.6 (76.8-93.4)	81 (74.7-87.2)	0.859 (0.802-0.917)
Radiologist 1	81.7 (71.6- 89.3)	72.8 (61.8- 82.1)	76.1 (65.8-84.5)	80(69.1-88.3)	77.3 (70.1- 83.4)	0.733 (0.698-0.847)
Radiologist 2	72.8 (61.8-82.1)	79 (68.5-87.3)	77.6 (66.6-86.4)	74.7 (64.2-83.4)	75.9 (68.6- 82.2)	0.755 (0.687-0.831)
			Mass+Thickening			
CNN	94.6 (81.8-99.3)	-	96.9 (84.2-99.9)	-	94.6 (81.8-99.3)	-
Radiologist 1	94.4 (81.3- 99.3)	-	97.1 (84.6-99.9)	-	94.4 (81.3- 99.3)	-
Radiologist 2	97.1 (85.1-99.9)	-	100 (90.5-100)	-	97.1 (85.1-99.9)	
			Polyp			
CNN	87.5 (47.3-99.6)	75 (34.9-96.8)	77.7 (39.9-97.1)	85.7 (42.1-99.6)	81.2 (54.3-95.9)	0.779 (0.529-0.994)
Radiologist 1	85.7 (42.1- 99.6)	62.5 (24.5-91.5)	80 (44.3-97.4)	85.7 (42.1-99.6)	73.3 (44.9-92.2)	0.759 (0.497-0.994)
Radiologist 2	75 (34.9-96.8)	75 (34.9-96.8)	85.7 (42.1-99.6)	77.7 (39.9-97.2)	75 (47.6- 92.7)	0.753 (0.497-0.994)
			Contracted			
CNN	93 (80.9-98.5)	71.4 (55.1-89.3)	78.7 (64.3-89.3)	57.5 (39.2-74.5)	84.5 (75.6-93)	0.860 (0.768-0.952)
Radiologist 1	81.4 (66.6- 91.6)	75 (55.1- 89.3)	83.3 (68.6-93.0)	72.4 (52.7-87.2)	78.9 (67.5- 87.6)	0.794 (0.680-0.907)
Radiologist 2	77.3 (62.2-88.5)	77.8 (57.7-91.3)	82.5 (67.2-92.6)	67.7 (48.6-83.3)	77.5 (66-86.5)	0.759 (0.640-0.877)

Learning from Limited Supervised Data

MICCAI 2023

Weakly Supervised GBC Detection

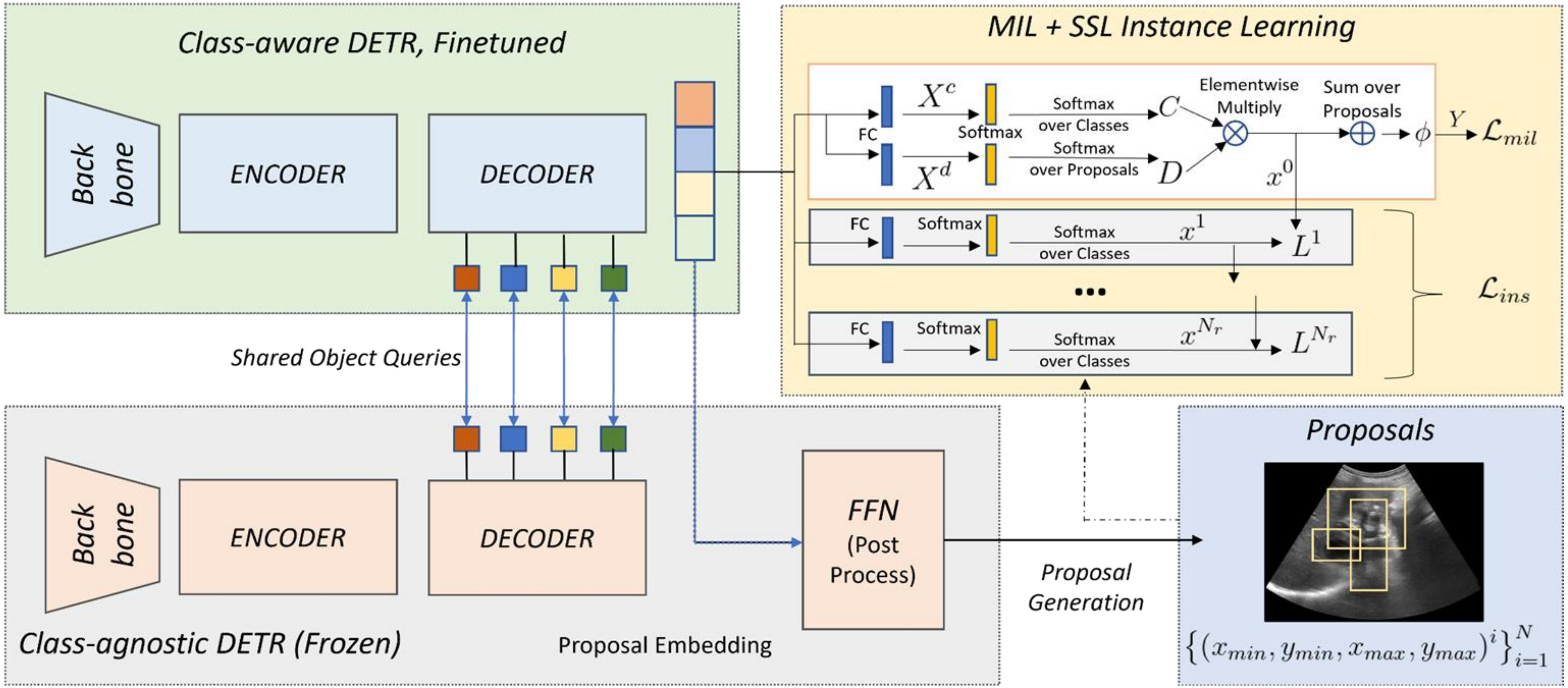
- Standard Image classifiers are difficult to train on GBC detection:
  - Low inter-class variance (a malignant region – small portion of a USG image),
  - High intra-class variance (sensor capturing 2D slice of 3D organ – large viewpoint variations)
  - Low training data availability

Weakly Supervised GBC Detection

- Training Object Detectors (bounding box output) – DNNs focus on the relevant ROI
- Bounding box annotations are costly to acquire
- Weakly Supervised Object Detection (WSOD) – train with only image labels
- Available without additional cost with diagnostic report
  - Eliminate the need of costly bounding box annotations – no additional effort from the human experts/ physicians

Weakly Supervised DETR

- Detection Transformer (DETR) is modified for weak supervision
- Novel DETR + MIL + SSL pipeline for generating bounding boxes



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Results

- Blue – GT box, Green – Predicted box
- Tested our model against 5 SOTA WSOD models
- Avg. Precision = 0.628 as compared to 0.531 by current SOTA (WS-DETR)