Expecto Petri-Dish

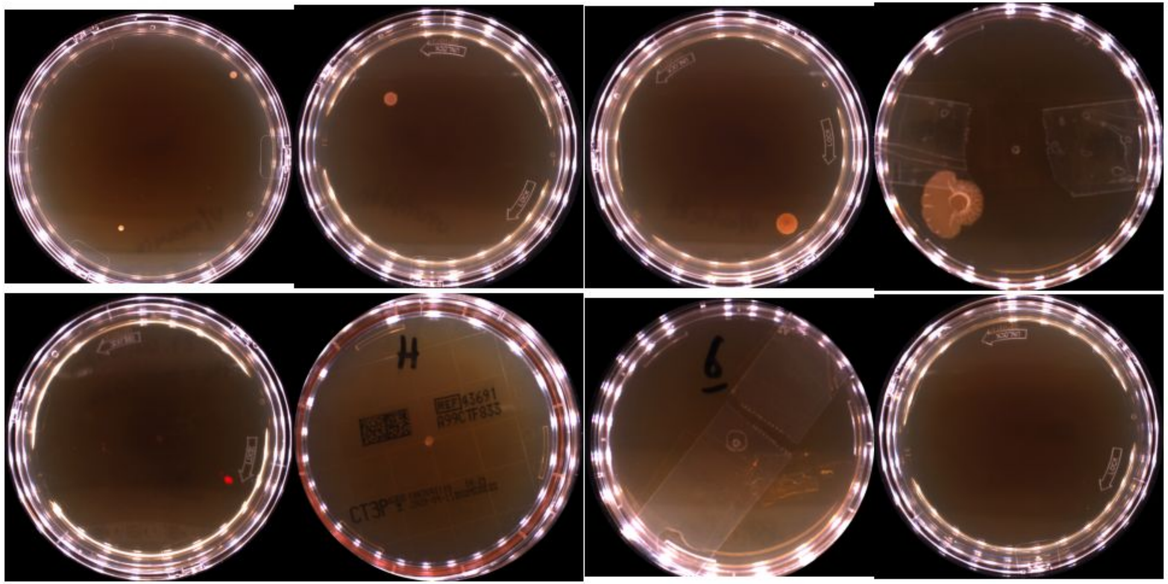
Team 2:
Anoop Raj Nair
Artem Lebedev
Sneha Narain

Berkeley
UNIVERSITY OF CALIFORNIA

Protein-based NN does sterility testing now. Silicon-based NN can be less biased.

When failing a sample is bad for business,
does that affect analyst's judgement?

Data Collection: 4500 expert labeled images



Contaminated

Sterile

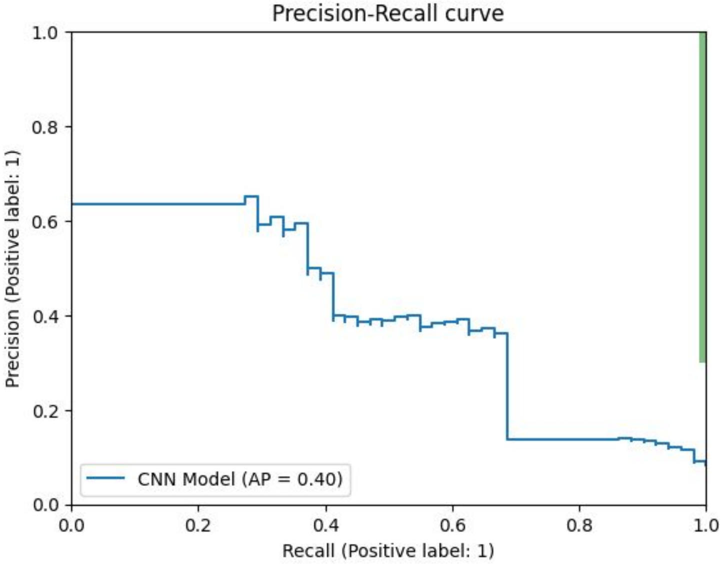
Recall over precision

Recall
Of all contaminated samples, how many did the model flag?

Precision
Of all the samples that the model flagged, how many were actually contaminated?

Target recall > 99.9%
be certain to find all contaminated samples

Target precision > 30%
It is Ok to include 2x as many non-contaminated



Block Diagram

Use Case
- Petri Dish Image Classification

Dataset Selection
- EDA
- Data Prep
- Train/Val/Test Splitting

Model Selection

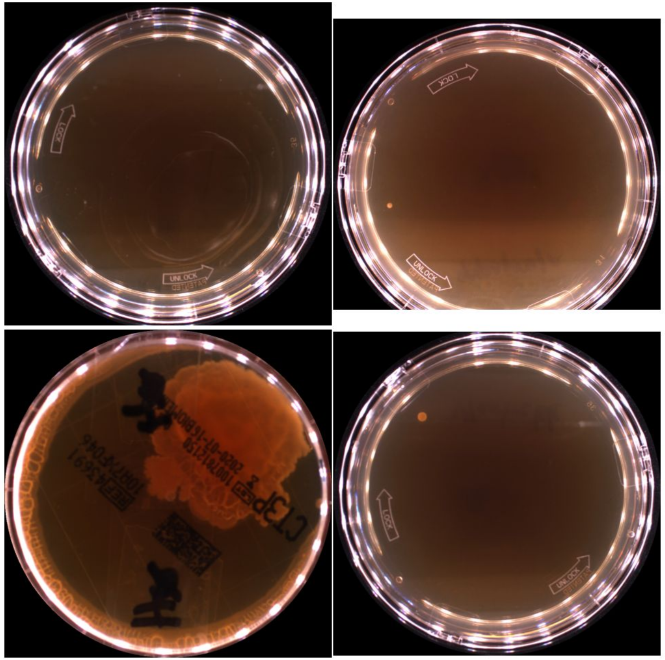
Classic ML:
- Random Forest
- CNN

Transfer learning:
- ResNet50
- VGG16
- Efficient Net
- ResNet50 Sequence 2
- Inception

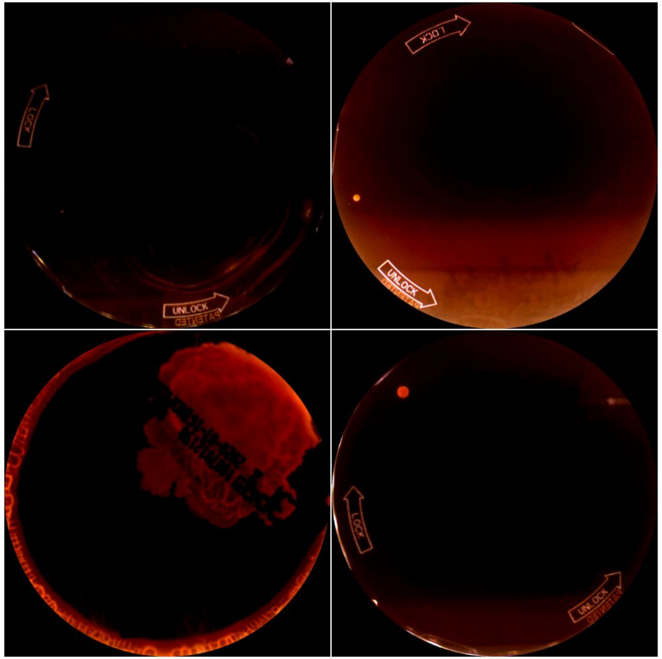
Evaluation
- F1 Score
- Accuracy
- Convergence
- Recall

Bias & Future Study
- Improved prep
- Data augmentation
- SAM + YOLO

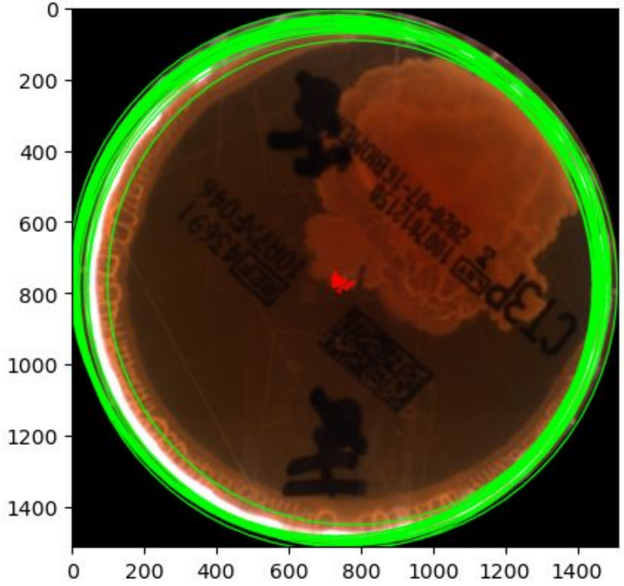
Image Processing



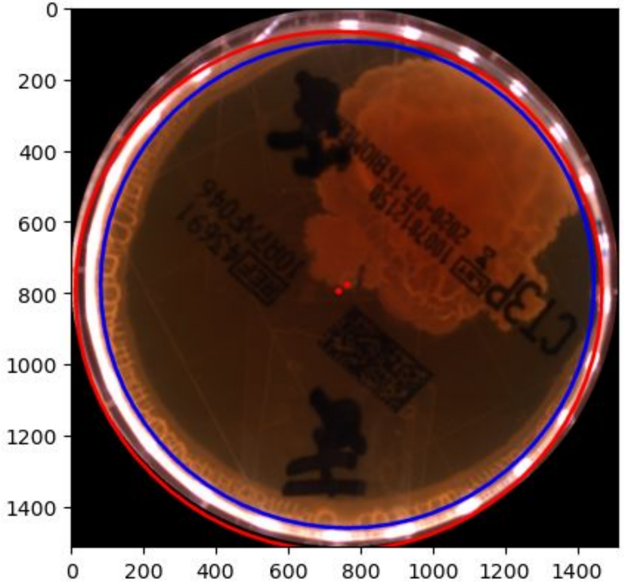
Hough-transform
Circular mask
Scale
Adaptive threshold
Contrast adjust



Pre-processing



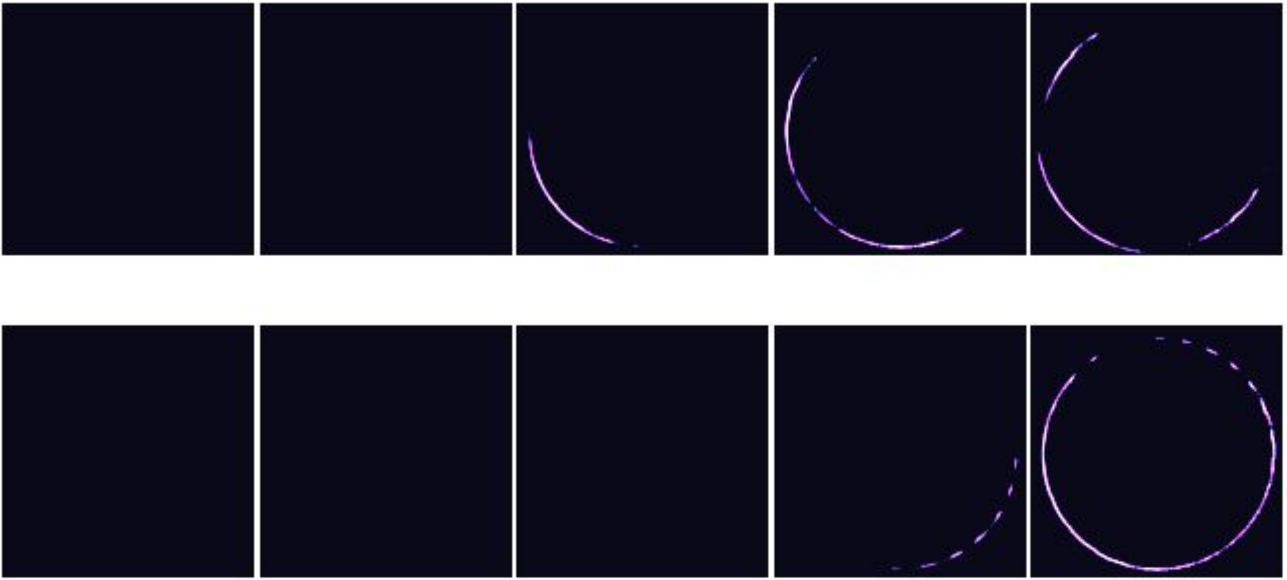
Circles suggested by
Hough-transform



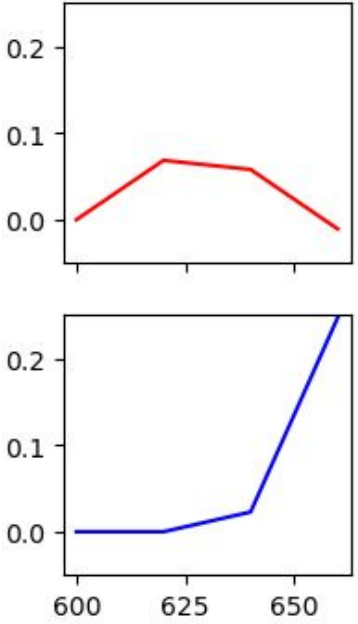
Best and worst
suggestions

Pre-processing

Intensity as we expand a thin ring centered where Hough suggested



Intensity derivative



Worst circle

Best circle

Data Splitting

Prior to subsetting the dataset

Original Dataset

~ Sourced from production environment

~ 4,352 total images

~ 3,743 images without the benchmark images

~ 315 positive images and 3,428 negative images

Subsetting the Dataset

~ 630 total images

~ Extracted 315 positive and 315 negative images from the original dataset

~ Does not use any of the images part of the Benchmark Dataset

Splitting the Dataset

~ Split dataset into training and validation

~ Trained the model using validation set size of 0.2

Benchmark Dataset

~ 14% of the original dataset

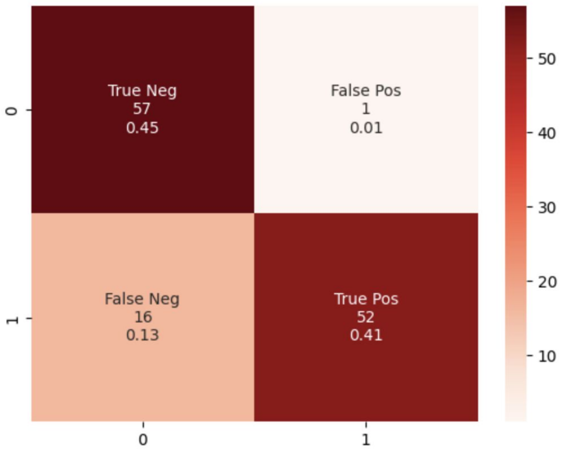
~ 609 total images

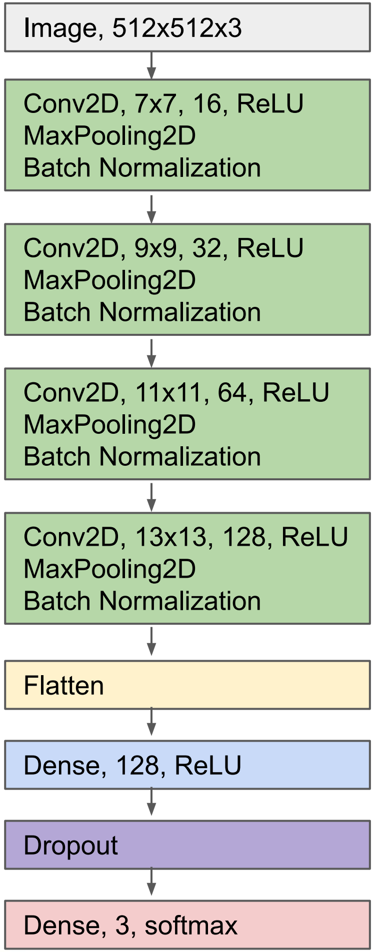
~ Images were not used to train or validate the model (only for testing)

Model Selection - CNN

~ Image resolution downsampled from 1500 to 512 pixels

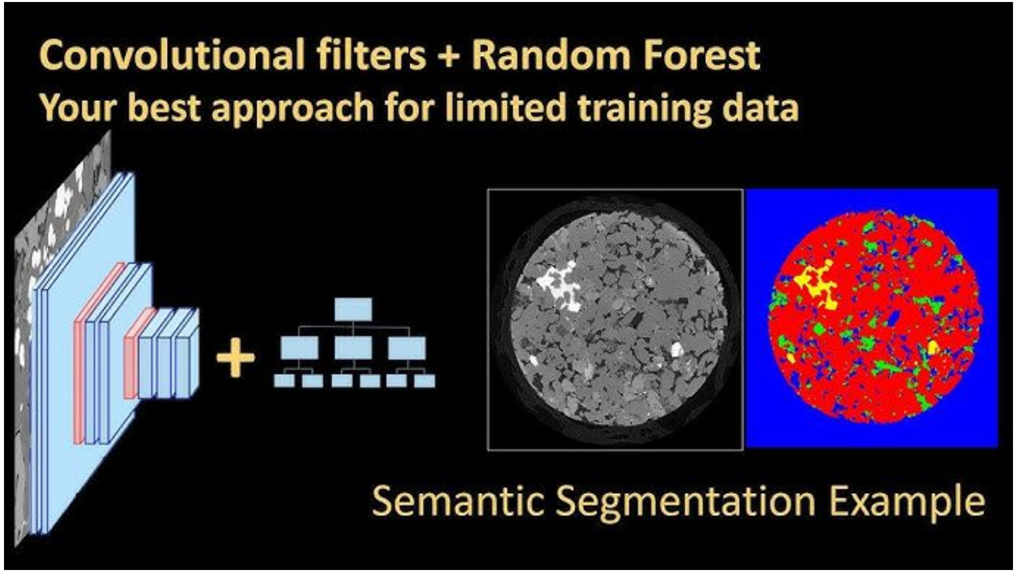
~ Incorporated 4 convolutional layers with varying kernel and filter sizes





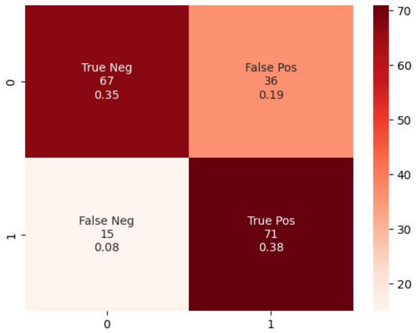
Berkeley
UNIVERSITY OF CALIFORNIA

Model Selection - Traditional

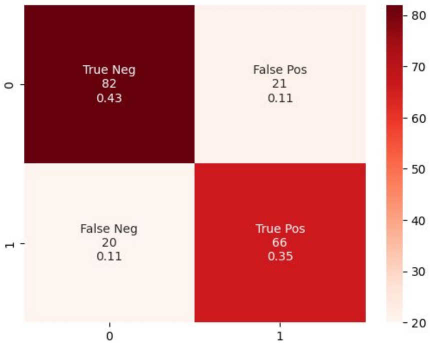


~700 training images: too few for CNN

On 128 x 128 images RF beats CNN



Traditional CNN Model

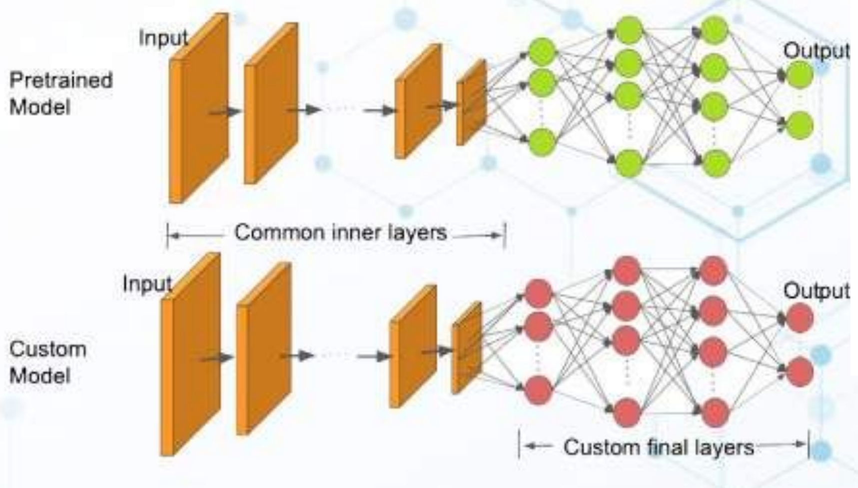


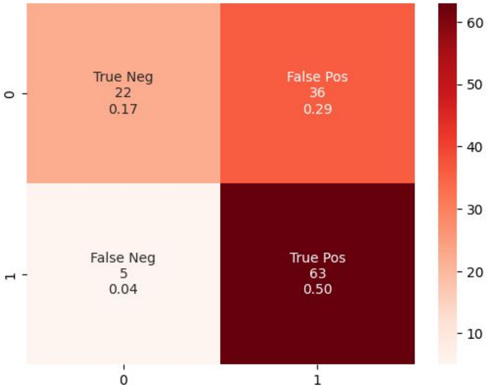
RF Model with Convolutional Filters

[Inspired by Digital Sreeni idea]

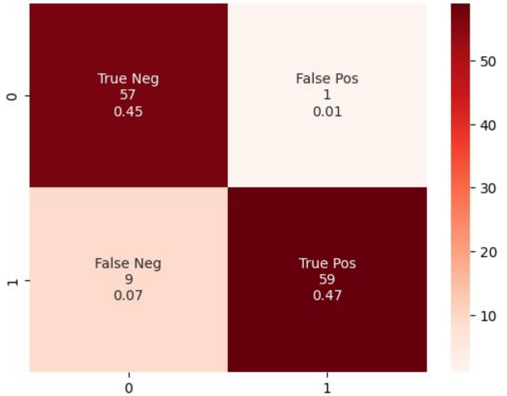
Berkeley
UNIVERSITY OF CALIFORNIA

Model Selection - Advanced





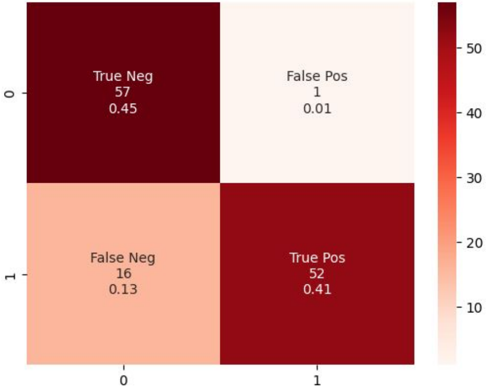
VGG16 Model



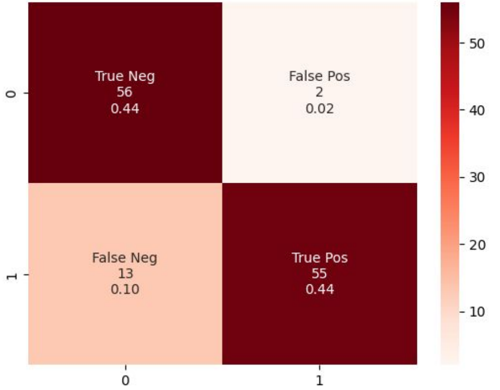
ResNet50 Model 2

~ Utilized the following pretrained models: VGG16, Efficient Net, ResNet50, and Inception

~ Efficient Net and ResNet50 Model 2 had the highest accuracies and F1 scores



Efficient Net Model



ResNet50 Model

Success/Failure Evaluation

ACCURACY

The proportional measure of the number of correct predictions over all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP *(true positives),* ***TN*** *(true negatives),* ***FP*** *(false positives),* ***FN*** *(false negatives)*

F1 SCORE

Measures the model's accuracy using its precision and recall scores. Good for models using imbalanced data.

$$F1\ score = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

CONVERGENCE TIME

Rate at which maximum or minimum of the model's loss/error function is reached

AVG PRECISION

Measures quality of the ranked list of predictions made by a model across different levels of recall.

Berkeley
UNIVERSITY OF CALIFORNIA

Model Comparison

	Training				Benchmark		
Model	Accuracy	F1 Score	Convergence	Average Precision	Accuracy	FI Score	Average Precision
Random Forest	0.80	0.80	0.18	0.79	0.79	0.83	0.71
CNN	0.86	0.86	0.30	0.94			
EfficientNet	0.87	0.87	0.18	0.96	0.95	0.95	0.77
ResNet50	0.88	0.88	0.17	0.98			
Resnet50 Seq2	0.85	0.85	0.08	0.98	0.93	0.94	0.75
VGG16	0.81	0.81	0.00	0.92	0.65	0.73	0.25
Inception	0.80	0.80	0.01	0.38			

Model Selection - Efficient Net

ResNet50-Seq2	Training	Benchmark
Accuracy	0.85	0.93
F1 Score	0.85	0.94
Average Precision	0.98	0.75
Convergence	0.08	

EfficientNet	Training	Benchmark
Accuracy	0.87	0.95
F1 Score	0.87	0.95
Average Precision	0.96	0.77
Convergence	0.18	

Criteria	ResNet	EfficientNet
Architecture	Based on residual learning with skip connections	Uses compound scaling for width, depth, and resolution
Advantages	- Easier training of deep networks - Good performance - Wide adoption and extensive research	- Better accuracy, - Better efficiency, - Scalability - Adaptable to different model sizes
Disadvantages	- May require more computational resources for deeper models - Not as efficient as EfficientNet	- Relatively newer architecture - Fewer pre-trained models available

Berkeley
UNIVERSITY OF CALIFORNIA

Ethics

NeurIPS 2021 Paper Checklist Guidelines

1	**For all Authors..**	**Response**
(a)	Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?	yes
(b)	Have you read the ethics review guidelines and ensured that your paper conforms to them?	yes
(c)	Did you discuss any potential negative societal impacts of your work?	yes
(d)	Did you describe the limitations of your work?	yes
2	**If you are including theoretical results...**	
(a)	Did you state the full set of assumptions of all theoretical results?	n/a
(b)	Did you include complete proofs of all theoretical results?	n/a
3	**If you ran experiments...**	
(a)	Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)?	yes
(b)	Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)?	yes
(c)	Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)?	no
(d)	Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)?	yes
4	**If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...**	
(a)	If your work uses existing assets, did you cite the creators?	yes
(b)	Did you mention the license of the assets?	yes
(c)	Did you include any new assets either in the supplemental material or as a URL?	no
(d)	Did you discuss whether and how consent was obtained from people whose data you're using/curating?	yes
(e)	Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content?	yes
5	**If you used crowdsourcing or conducted research with human subjects..**	
(a)	Did you include the full text of instructions given to participants and screenshots, if applicable?	n/a
(b)	Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable?	n/a
(c)	Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation?	n/a

Future Work

~ Gather more data/images from other production environments

~ Identify the number of bacterial colonies on a petri dish using **SAM + YOLO**

~ Fine tune the Efficient Net and ResNet models

…

Berkeley
UNIVERSITY OF CALIFORNIA

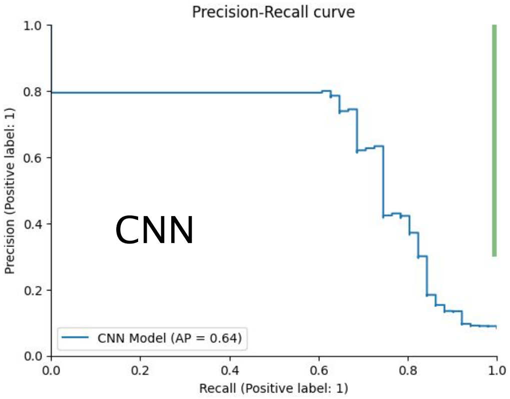
Backup Slides

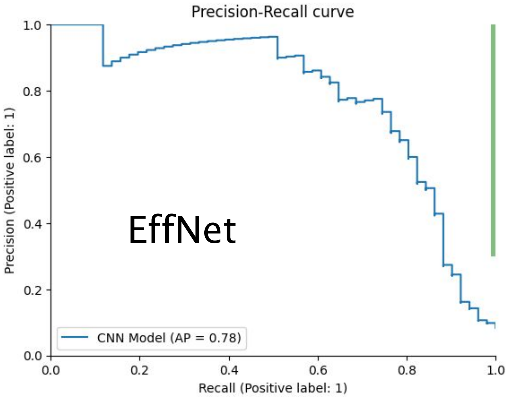
Table Comparison (Validation)

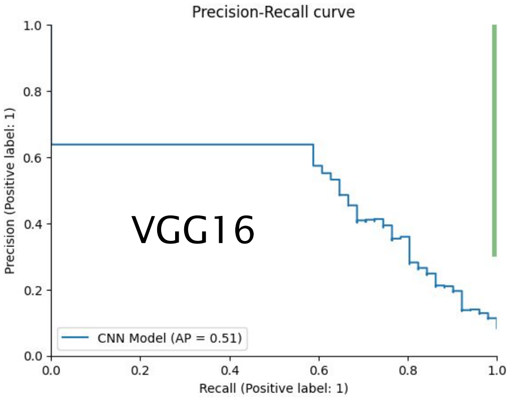
Model	Random Forest	CNN	EfficientNet	ResNet50	VGG16	Resnet50 Seq2	Inception
Accuracy	0.80	0.86	0.87	0.88	0.81	0.85	0.80
F1 Score	0.80	0.86	0.87	0.88	0.81	0.85	0.80
Convergence	0.18	0.79	0.18	0.17	0.00	0.08	0.01
Average Precision	0.79	0.94	0.96	0.98	0.92	0.98	0.38

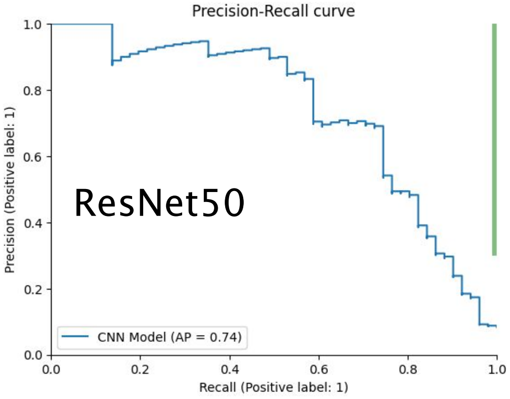
Table Comparison (Benchmark)

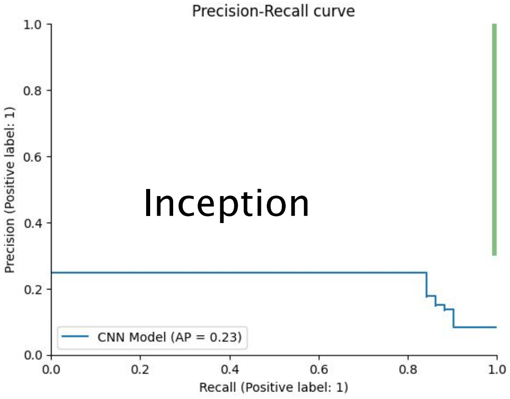
Model	Random Forest	CNN	EfficientNet	ResNet50	VGG16	Resnet50 Seq2	Inception
Accuracy	0.79		0.95		0.65	0.93	
FI Score	0.83		0.95		0.73	0.94	
Average Precision	0.71		0.77		0.25	0.75	

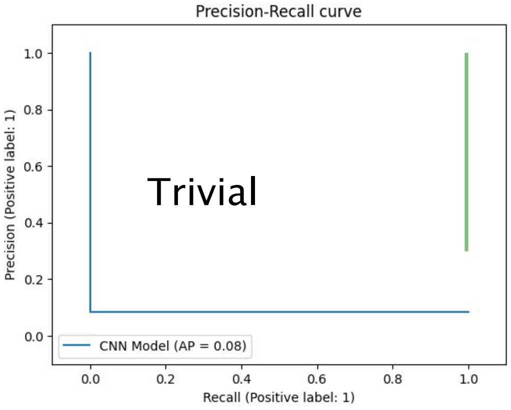

Precision-Recall curve
CNN
CNN Model (AP = 0.64)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall curve
EffNet
CNN Model (AP = 0.78)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall curve
VGG16
CNN Model (AP = 0.51)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall curve
ResNet50
CNN Model (AP = 0.74)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall curve
Inception
CNN Model (AP = 0.23)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall curve
Trivial
CNN Model (AP = 0.08)
Precision (Positive label: 1)
Recall (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0

Fairness

Demographic Fairness	Not applicable
Equal Opportunity	We computed **TPR** (true positive rate) and there are no significant differences among groups to indicate unequal opportunity
Model Transparency	We can enhance the deep learning models to include techniques like **LIME** (Local Interpretable Model-agnostic Explanations) or **SHAP** (SHapley Additive exPlanations) to generate explanations for individual predictions.
Model Robustness	Our dataset is limited which limits data augmentation. However, we created split datasets to ensure model training and validation on appropriate samples and then compared model performance against benchmark set.

Berkeley
UNIVERSITY OF CALIFORNIA