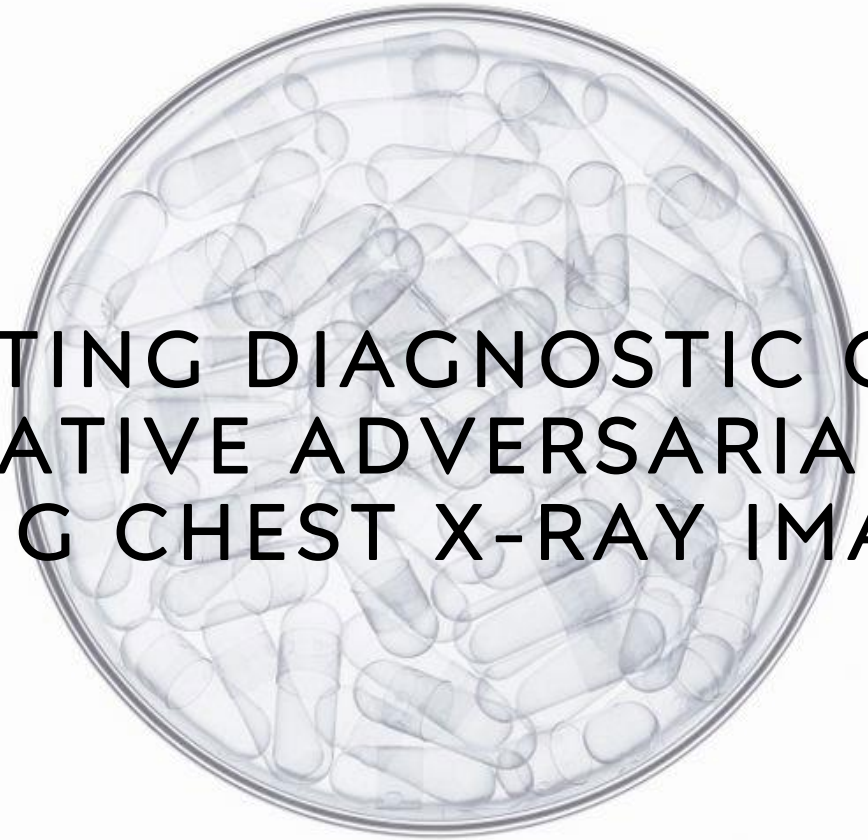
SUBSTANTIATING DIAGNOSTIC CAPABILITIES WITH GENERATIVE ADVERSARIAL NETWORKS USING CHEST X-RAY IMAGES



ABHISHEK SRIVASTAVA

AGENDA

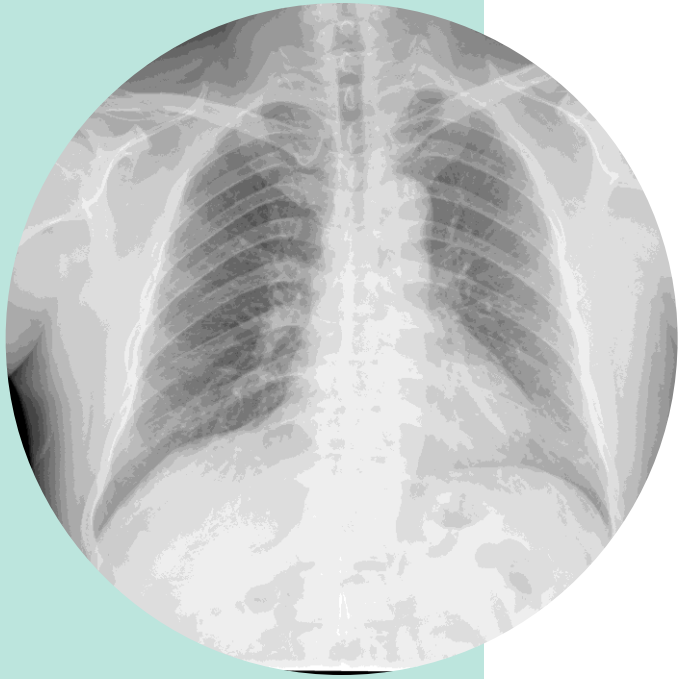
MOTIVATION

APPROACH

DATASET

RESULTS

CHALLENGES



NEARLY 3 MILLION TUBERCULOSIS CASES REMAIN UNDIAGNOSED EACH YEAR

Problem

Early and accurate diagnosis of Tuberculosis

Solution

Automated and real-time prediction of Tuberculosis with certain confidence

Why GAN?

Limited unsupervised disease detection models available

It is more difficult to find examples of tuberculosis patients than healthy patients

Can be extended for diagnosis of other pulmonary diseases

AGENDA

MOTIVATION

DATASET

APPROACH

RESULTS

CHALLENGES

TRAINING AND TESTING ON MULTIPLE DATA SOURCES ADDS GENERALIZABILITY

AGENDA

MOTIVATION

DATASET

APPROACH

RESULTS

CHALLENGES

GENERATE HEALTHY IMAGE PATTERNS USING GAN

Data Pre-processing

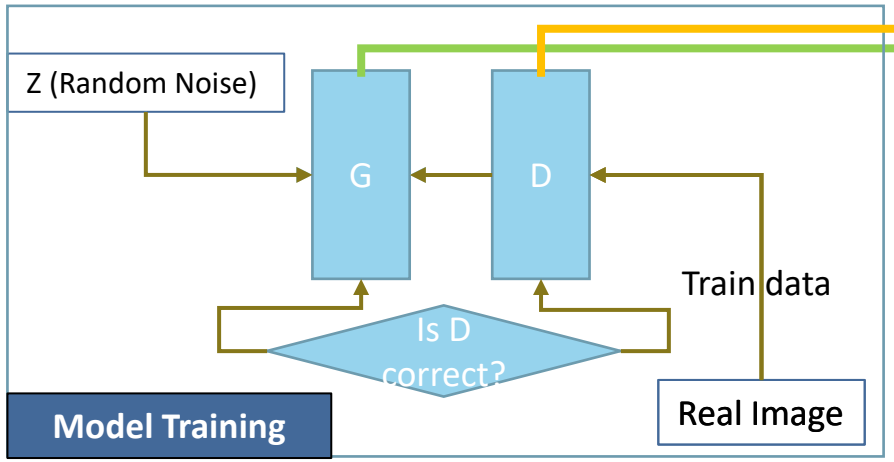
- Pre-process chest X-ray images

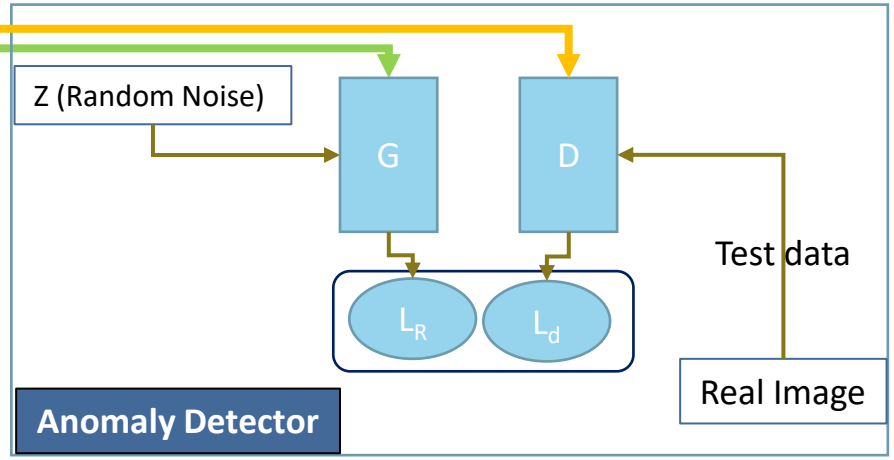
Model training

- AnoGAN Architecture for Anomaly Detection
- Train AnoGAN on Healthy 'original' images

Optimization and App development

- Identify appropriate threshold of anomaly score
- A Streamlit app predict Tuberculosis provided an X-Ray images





CHOOSE THRESHOLD TO DETECT ANOMALOUS IMAGES

The GAN model attempts to generate real-like (non-anomalous or healthy) images, while generator and discriminant follows two-player minmax game with value function V(G,D)

$$\min_G \max_D V(D,G)$$

$$V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z))]$$

When GAN is supplied with a test anomalous image, total anomaly score is calculated which is a weighted sum of Residual Loss, L_R (λ) and Discriminant loss, L_D (1 - λ)

$$\mathcal{L}_R(\mathbf{z}_\gamma) = \sum |\mathbf{x} - G(\mathbf{z}_\gamma)|$$

$$\mathcal{L}_D(\mathbf{z}_\gamma) = \sum |\mathbf{f}(\mathbf{x}) - \mathbf{f}(G(\mathbf{z}_\gamma))|$$

The distribution of anomaly scores from a well trained model should lie at a separable distance for healthy and un-healthy patients

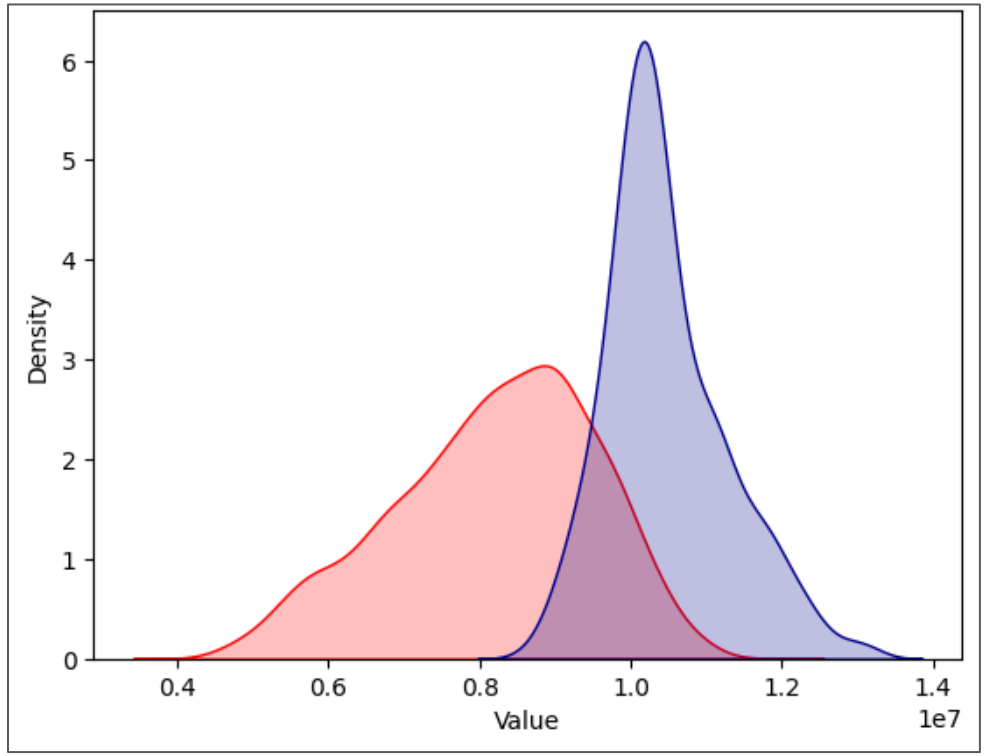
TBX11K AND SHENZHEN HEALTHY X-RAYS WAS USED TO TRAIN ANOGAN

Training Data			
Patients	**TBX11K**	**Shenzhen**	**Montgomery County**
Sick or Tb	0	0	0
Healthy	3,377	251	0
Validation Data (To decide threshold)			
Tb	101	201	0
Testing Data			
Healthy	348	75	80
Tb	699	135	50
Sick and Non-Tb	3,800	0	0

Note
Montgomery was not used while training the GAN, to evaluate performance of GAN with unseen chest X-Rays images from different sources.

**Tb stands for Tuberculosis*

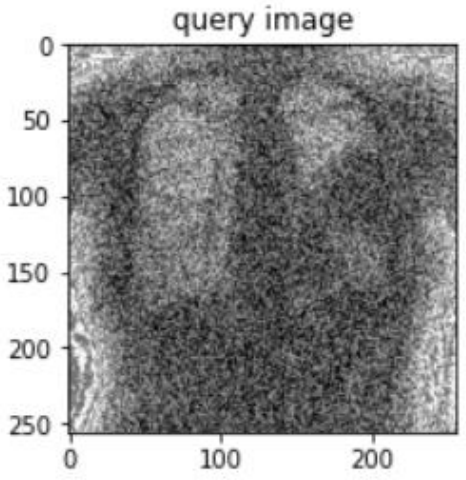
ANOMALY SCORES WERE OBSERVED ON VALIDATION SET TO CHOOSE THRESHOLD

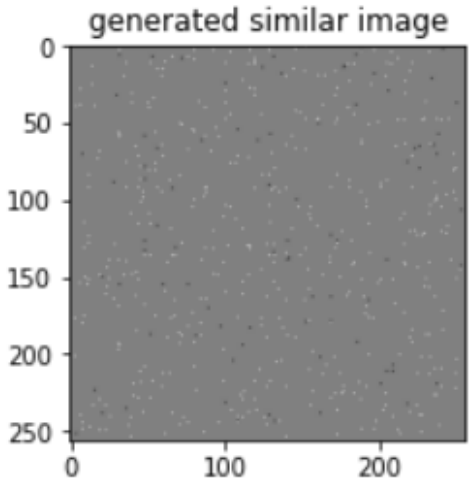


Healthy (Red) and Tb (Blue) anomaly scores on validation set after hyper parameter tuning

Anomaly score of 9.5 x 10^6 was chosen as the threshold to best classify healthy and Tuberculosis patients

While training GAN, the attempt is to train generator to reconstruct similar discernible healthy images however, in case of limited computational power, GAN may learn to reconstruct healthy pattern just enough enabling detection of desired diagnosis





PARAMETERS WERE TUNED FOR BEST SEPARATED HEALTHY AND TB PATIENTS

Hyper-parameters	Explored	Used
Length of Noise vector (Z)	[10, 20, 50]	10
Image dimensions	28X28; 64X64; 256X256	256X256
Proportion of discriminator and residual loss	λ = [0.5, 0.9, 1]	0.9
Optimizer	Adams, RMSprop	RMSprop
Learning rate	For discriminator = [0.0001, 0.004] For generator = [0.0001, 0.002]	For discriminator = 0.0004 For generator = 0.0002

AGENDA

MOTIVATION

DATASET

APPROACH

RESULTS

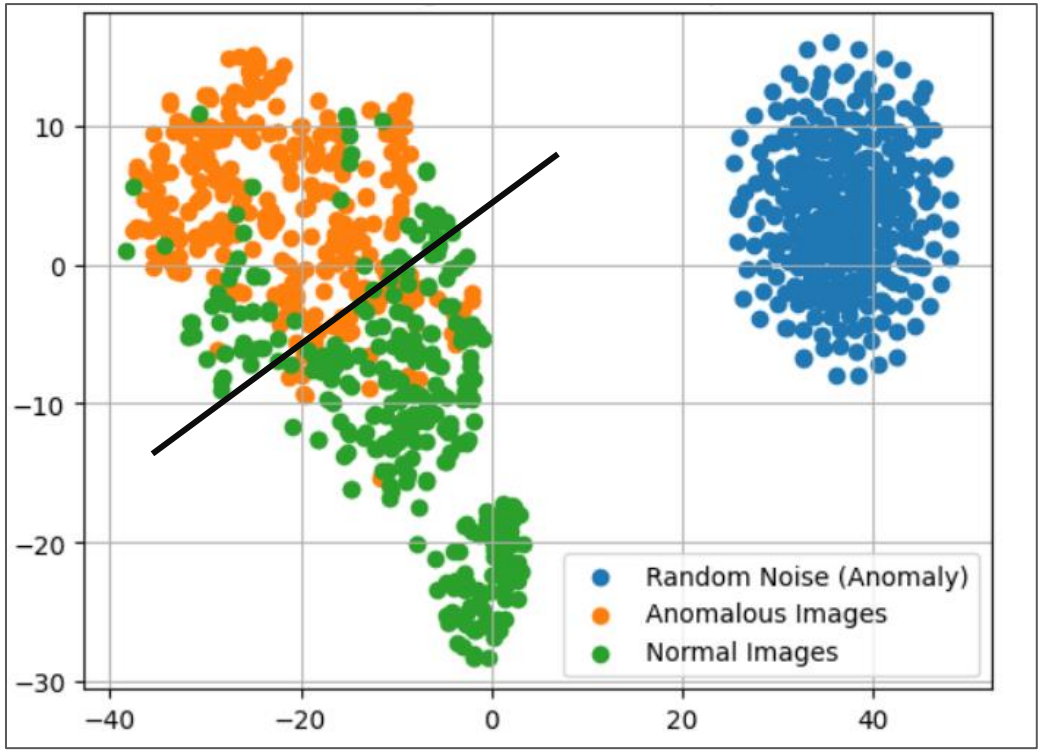
CHALLENGES

MODEL PERFORMED BEST WITH SHENZHEN AND WORST WITH MONTGOMERY DATA

Model Performance on Test data		
Datasets	Accuracy	F1-Score
TBX11K	71.5%	65.8%
Shenzhen	84.7%	85.7%
Montgomery	55.1%	41.2%

1. *Highest recall of 94.1% was observed on Shenzhen dataset, it only included pulmonary Tb patients*
2. *Lowest model accuracy and F1-score on Montgomery dataset concludes that the GAN trained is subject to source of Chest X-Ray images*
3. *On sick and Non-Tb, an accuracy of marginally more than 50%, which could be because such sick patients may not have a pulmonary disease*

GAN LEARNT SENSIBLE ANATOMICAL FEATURES



t-SNE embedding of normal (green) and anomalous (orange) images on the feature representation of the last convolution layer of the discriminator.

1. *Almost well separated normal and anomalous images showcased that the discriminator was able to learn features while training that separates healthy and sick patients*

2. *Embedding of random noise was clearly far away from normal and anomalous images, highlighting that the model can not be fooled when supplied with normal noise*

AGENDA

MOTIVATION

DATASET

APPROACH

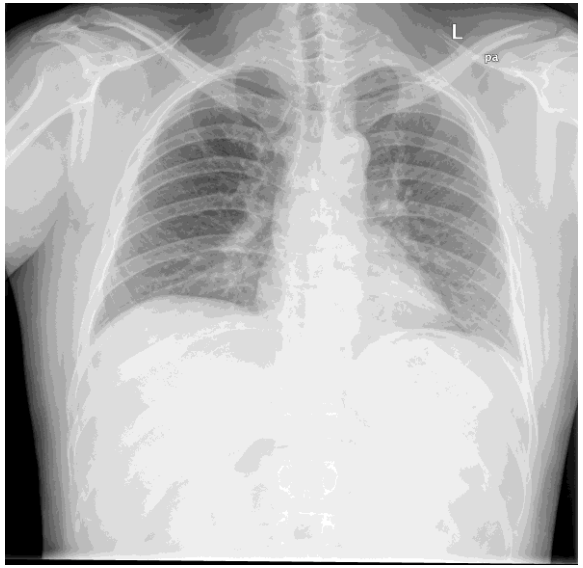
RESULTS

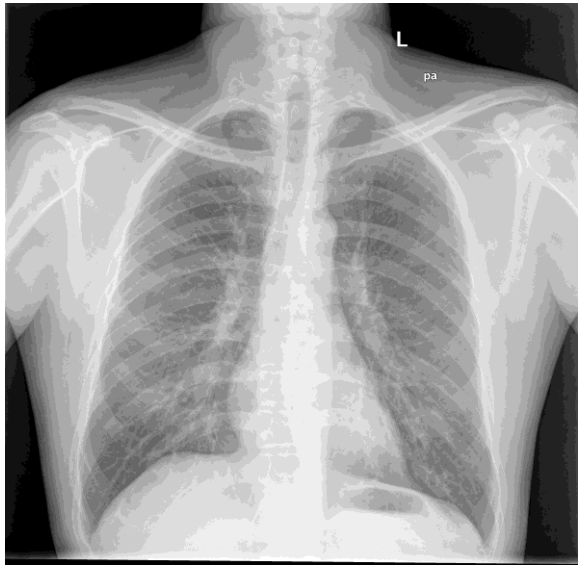
CHALLENGES

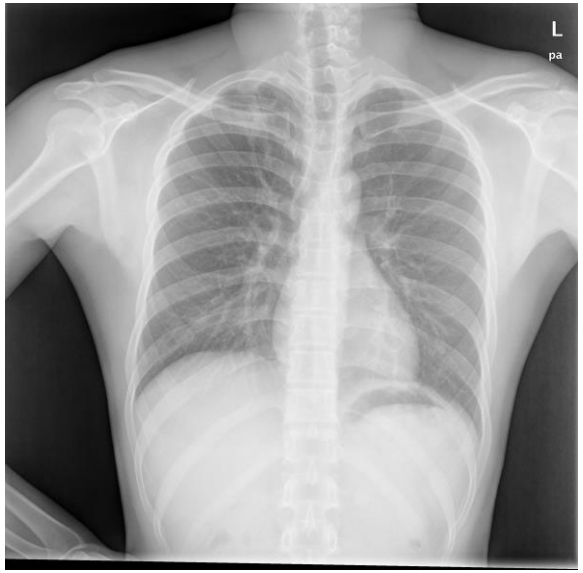
GANs ARE COMPUTAIONALLY EXPENSIVE

- Current AnoGAN architecture takes > 2 hour to train on 3,800 images of shape 256X256 from TBX11k and Shenzhen data
  - Find detailed architecture and code on my GitHUB
- Model deployment using Streamlit takes 3-4 minutes for a single prediction
- The current architecture deployment would not recommend predicting tuberculosis for test image with size less than 256 X 256
- GANs performance for this use case was strongly found subject to source of chest X-Ray images, resulting in poor performance on Montgomery dataset

QUESTIONS?

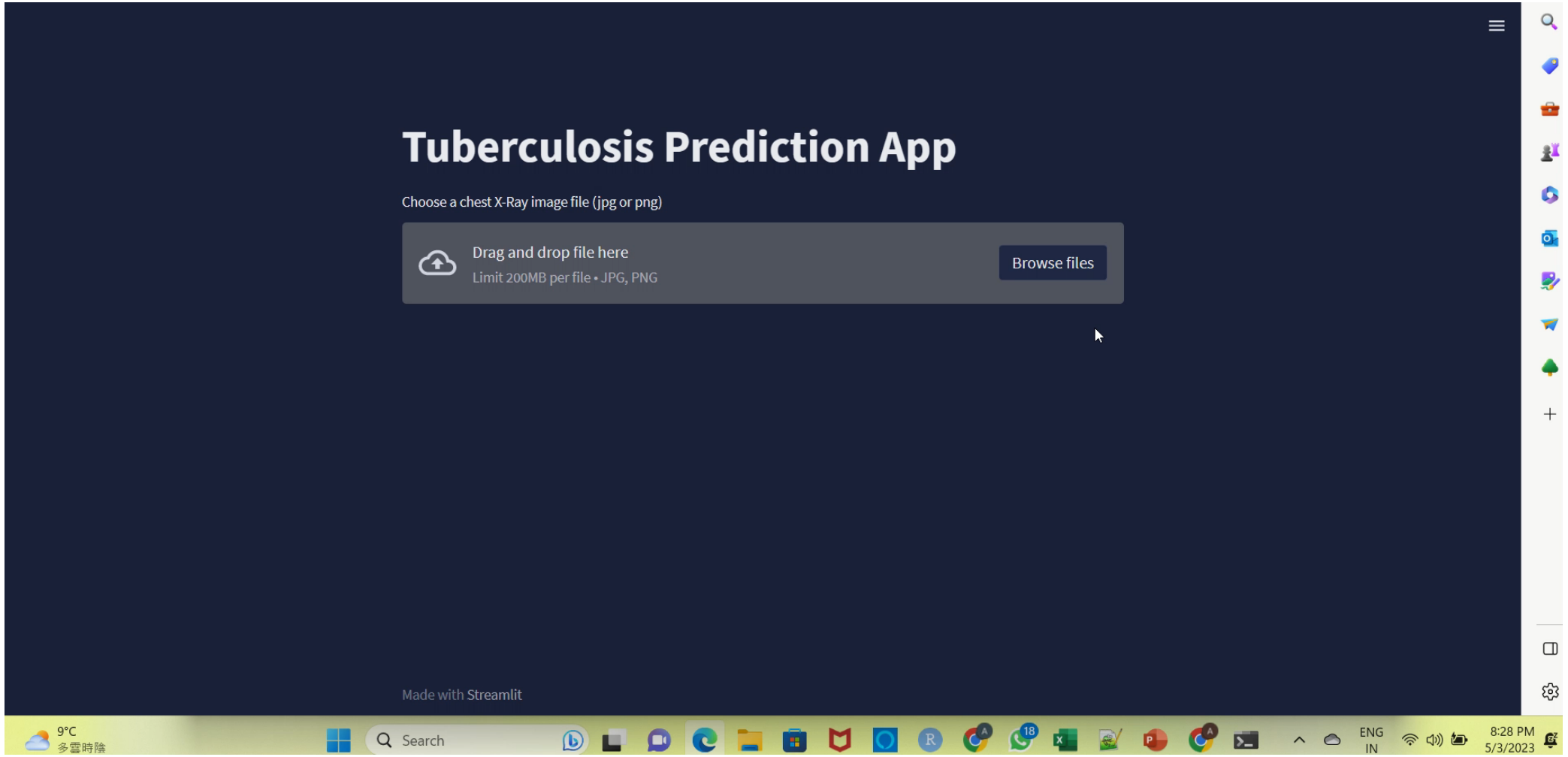


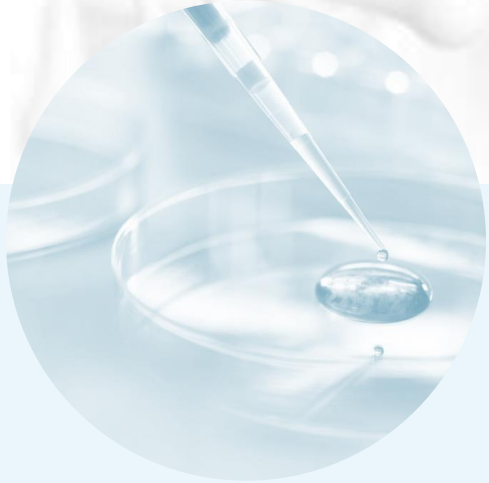




Which one has Tuberculosis, Can you tell?
Let's see if the "Tuberculosis prediction App" can find out!

STREAMLIT APP DEMO





THANK YOU

ABHISHEK SRIVASTAVA