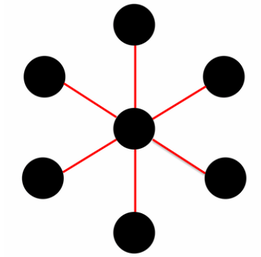
NaN.ai



A general purpose AI and machine learning startup leveraging federated learning to perform analytics on data that you can't see

Natan Vidra: vidranatan@gmail.com

Situation

There exists many organizations that would like to leverage AI and machine learning to make predictions, perform analytics and extract insights. There is more data out there than ever before, sitting in individual silos specific to these organizations.

Complication

Though many organizations would benefit from cooperating with one another, they are unable to share data across different silos due to legal and privacy constraints. As a result, many organizations implement models based on data local to their specific silo, which results in lower performance and less robust models.

Data Silos in Hospitals

Hospital A is looking at images of mammograms to detect whether or not someone has breast cancer. They implement a classical machine learning model on mammogram images. While the model performs well, the accuracy is not as great as they expected. If only this hospital had access to more data, then the machine learning model would perform better.

They realize that there are many of different hospitals out there, such as hospital B and C, each with very similar mammogram image data. However, due to legal and privacy constraints related to sharing data, they are forced to stick to their siloed data source for model analysis.

A

B

C

Breast Cancer Globally in 2020

Still a massive problem:

2,260,000 incidents per year

685,000 deaths per year

1 in 8 women affected per year

Lots of image data out there:

880,000,000 Mammograms/year

But very few publicly available, centralized datasets:

20 known public research datasets

<.1% of data is publicly available

Data is biased, solely from the states of NY, MA, & CA

What do people currently do to detect breast cancer?

Option 1: Get a Radiologist or Specialist to look at specific Mammogram Images

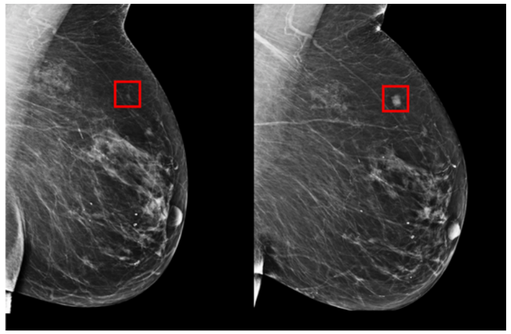
1/4 false positive/negative rate for radiologist

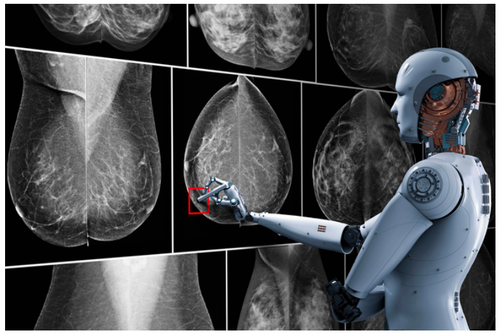
Radiologists are extremely expensive

Option 2: Train Classical Machine Learning Models on Local Data

1/3 false positive/negative rate for radiologist

Local data leads to model bias & non-generalizable performance



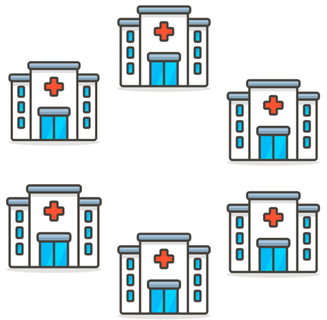


Why can't people centralize all data across different hospitals?

Option 3: Centralized Data for more global classical machine learning models in a central server

Cooperating to get better AI models to predict breast cancer would be great, but due to HIPPA compliance hospitals can't share Protected Health Information (PHI) and Personal Identifiable Data (PID), such as mammograms.

Exponential amount of legal work for each hospital that wants to add de-anonymization to data to build centralized medical models.



If there are n = 6 hospitals, each hospital needs n-1 = 5 legal agreements with other hospitals.

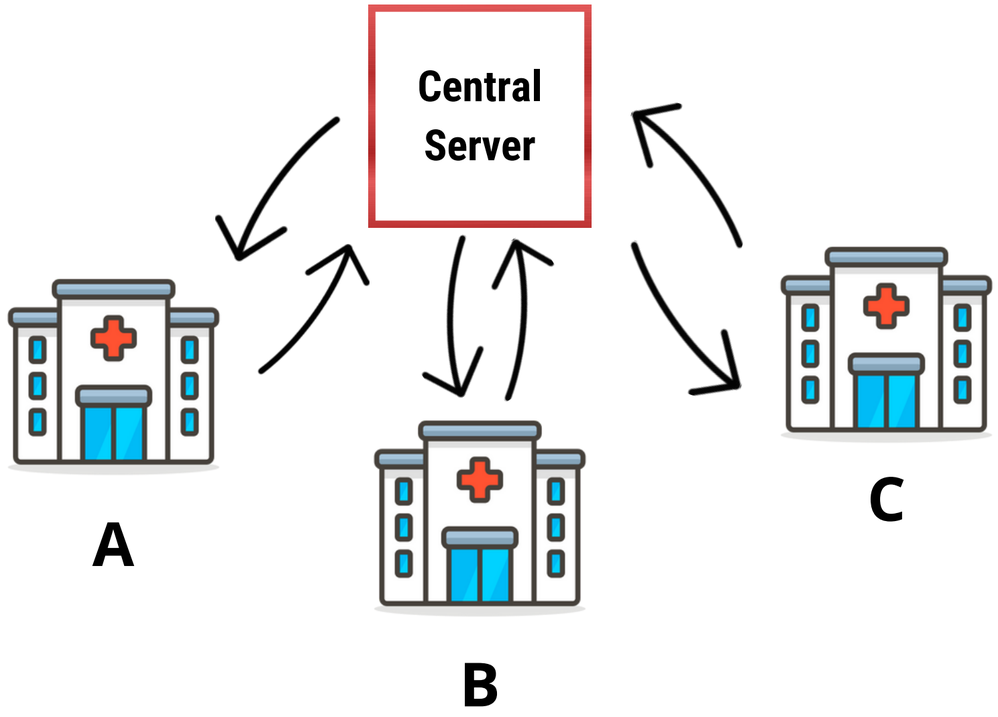
This translates to O((n-1)^2) agreements as n becomes large, which is not scalable

NIH
National Institutes of Health

It can take organizations such as the National Institute of Health 6 months to a year to centralize a large enough, de-anonymized/ unbiased dataset ready for AI model analysis

What is Federated Learning?

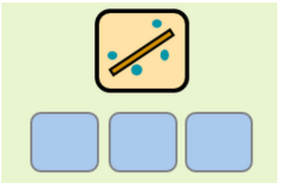
Federated learning is a privacy preserving approach to training machine learning models across a decentralized network of data providers



How does Federated Learning Work?

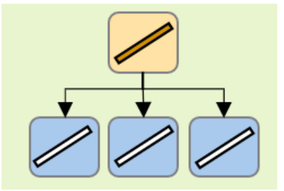
1) Initialize Global Model

A central server creates an initial model using only locally available data



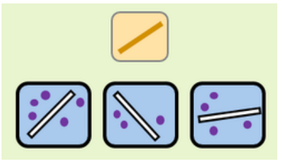
2) Broadcast Global Model

Initial model is sent to each data provider



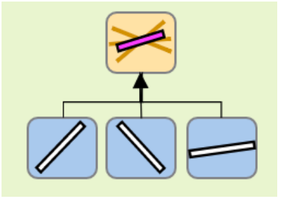
3) Retrain Locally

The data providers update the model by training on their own local data



4) Update Global Model

The model updates *only* are sent back to the central server and aggregated into an improved model



How can you make money from this?

Data Centric

Provide hospital A's models access to hospital B's and C's data sources *(images of mammograms)* for a fixed or variable cost

Model Centric

Provide hospital A access to enhanced modeling capabilities for predicting breast cancer from *images of mammograms*. Hospital A can now generate more accurate insights and reduce model bias, for a fixed or variable cost

People are starting to use Federated Learning to predict breast cancer today

NVIDIA

Flywheel Exchange
American College of Radiology
Rhino Health
Taiwan Web Service Corporation

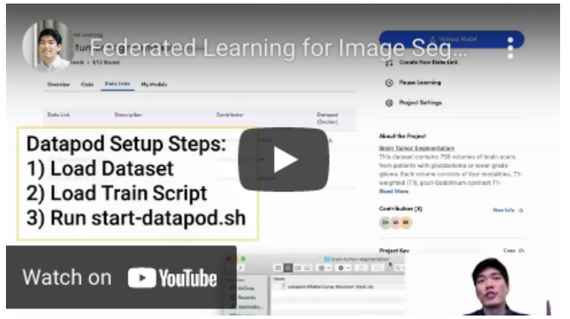
Healthcare Startups

Owkin
Apheris
Arkhn
Sharecare / Doc.ai
Lifebit
SAIL
DynamoFL

Non-Profits

PySyft
National Institute of Health

Rhino Health is using NVIDIA Flare to do federated learning for medical imaging p in the Massachusetts General Hospital

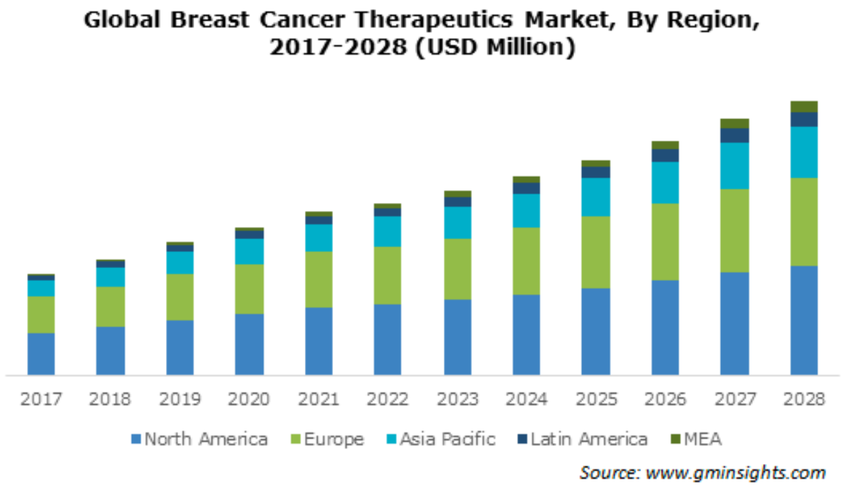


DynamoFL's low code federated learning solution is being deployed in hospitals in Massachusetts, like the Brigham and Womens Hospital, today

Market Opportunity?

What is the Total Addressable Market (TAM)?

- $22.5B TAM in 2021
- CAGR of 8%
- Projected TAM of $42.6B in 2028

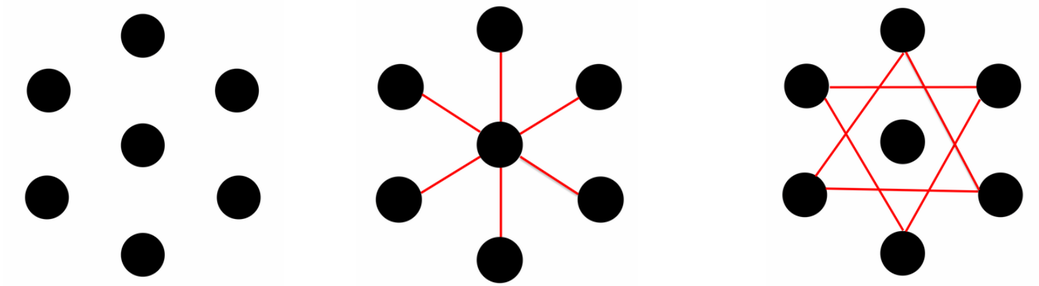


Who are the key players?

AstraZeneca
Pfizer
Novartis
Merck
Fresenius
Eli Lilly

What is the vision of NaN.ai?

A general purpose AI and Machine Learning startup leveraging federated learning to perform analytics on data that you can't see



Classical ML Federated ML Fully Decentralized ML

And here is a bit about me: Natan Vidra (vidranatan@gmail.com)

Data Scientist at Deloitte from 2021 - 2022

Founder and CEO of Flamingle from 2020 - 2021

M.Eng. in Computer Science, B.S. in Electrical and Computer Engineering from Cornell from 2015 - 2020