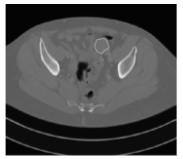
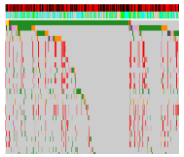



Radiogenomics investigation in metastatic breast cancer using data from oncology clinical trials

Craig Wang on behalf of the FDA-Novartis Radiogenomics Collaboration*

Basel Biometric Society (BBS) Seminar
6th November 2023, Roche, Basel

*Consisting of FDA OCE, CBER, CDER, CDRH, and Novartis, see here for list of contributors

NOVARTIS | Reimagining Medicine

Disclaimer

This presentation reflects views of the presenter and should not be interpreted as the official views of Novartis or FDA.

Introduction

Cancer is a complex and devastating disease that impacts millions of lives worldwide

Caused by abnormal cell growth which invade other parts of body

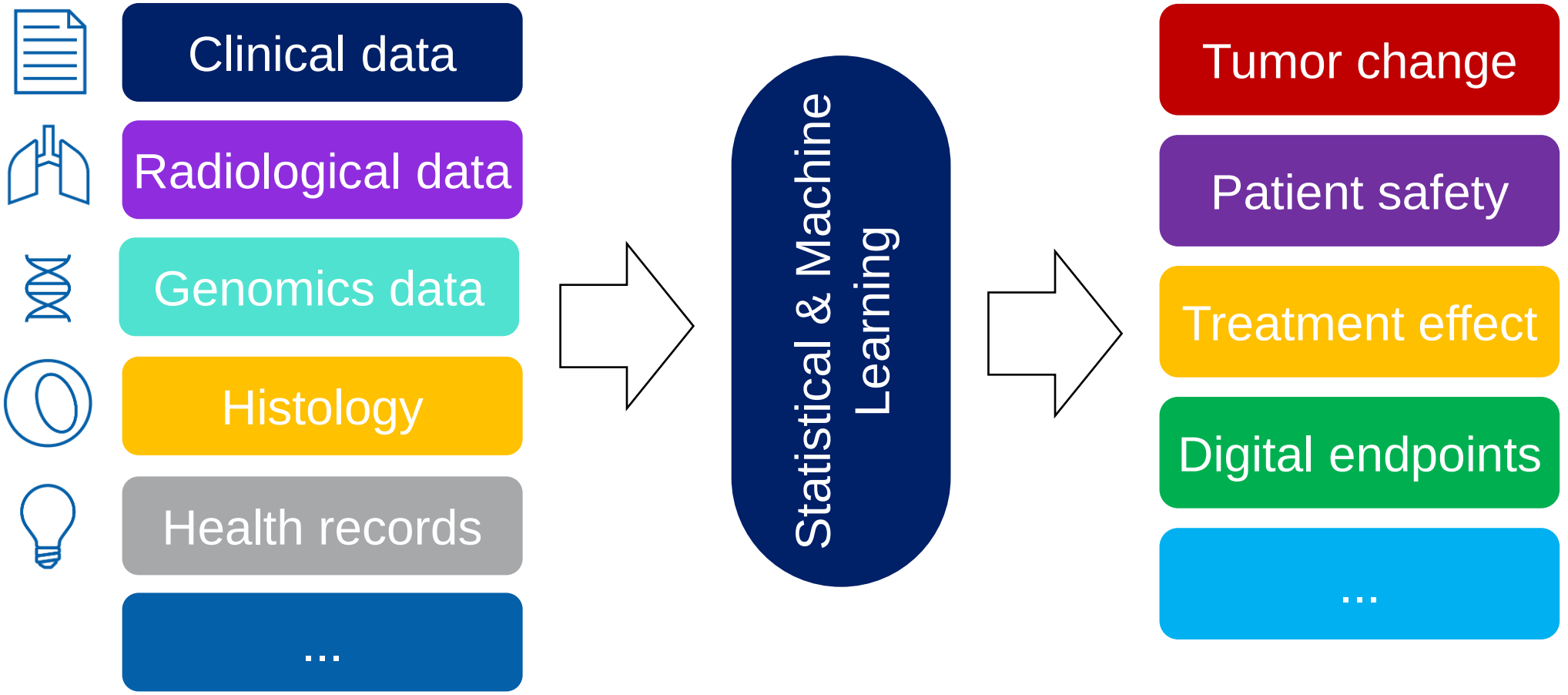
Involving both microscopic and macroscopic changes yet to fully understood

Oncology studies aims to advance the knowledge of cancer, its causes (prognostic), prevention, diagnosis and treatment options (predictive)

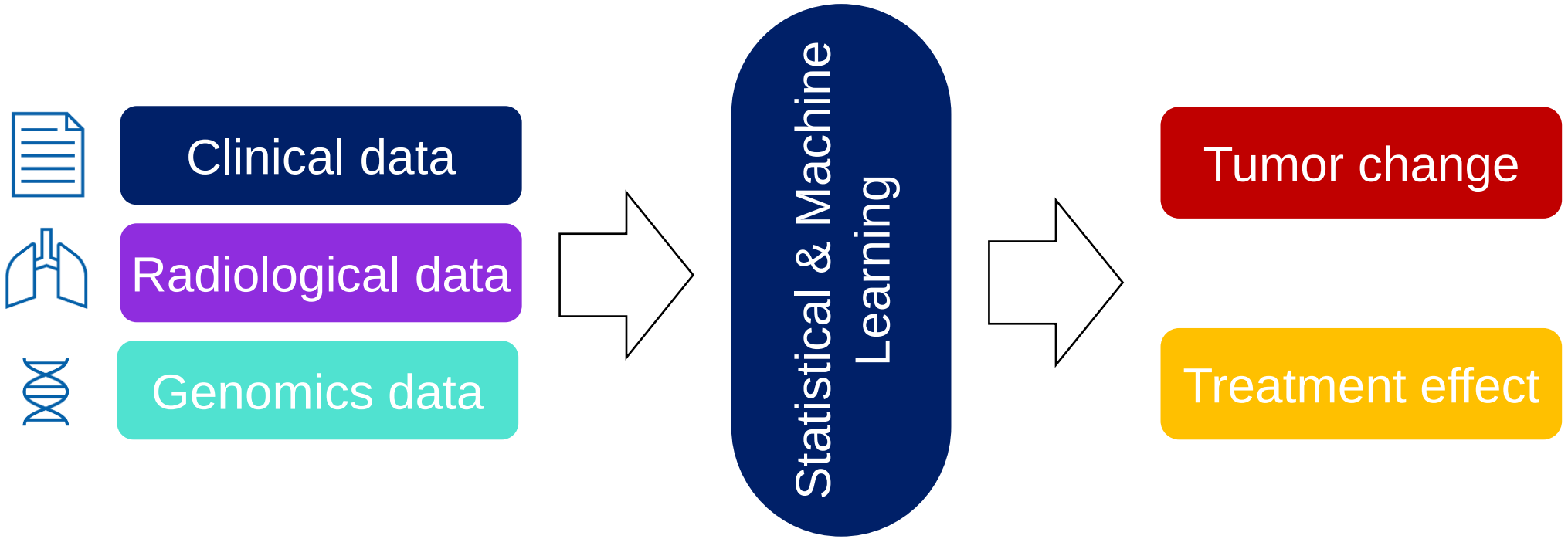
New cancer treatments aim to delay tumor progression and ultimately prolongs survival

Deeper interrogation of data potentially allows more targeted treatments, minimizing side effects and optimizing patient outcomes

Emerging data modalities in Oncology



Emerging data modalities in Oncology



Collaboration background

In 2020, Novartis and FDA started a scientific collaboration

Approach complex new data modalities and advanced analytics to interrogate clinical, radiomics and genomics data from clinical trials

The collaboration uses anonymized data from three Phase III trials sponsored by Novartis which investigates the drug ribociclib

Ribociclib is a targeted therapy used in combination with hormonal therapy shown to prolong both progression-free survival (PFS) and overall survival (OS) in patients with metastatic breast cancer

The top scientific goals is to evaluate and identify novel clinical and radiogenomics-based prognostic and predictive markers to advance public health

Additionally, the collaboration provides insights into research projects in machine learning and AI

Metastatic breast cancer

In 2020, 2.3 million women diagnosed with breast cancer and 685 000 deaths globally [1]

Majority of death caused by metastatic breast cancer (mBC)

HR+ HER2- mBC is a sub-type of breast cancer

Treatments for these patients typically include endocrine combination therapy (such as Ribociclib + Fulvestrant) or chemotherapy

In mBC trials, the primary endpoint is typically PFS based on RECIST 1.1 criteria, with a secondary endpoint OS

Disease progression status is detemined by the change in sum of lesion diameters, max. 5 target lesions are identified at baseline to track them overtime

[1] https://www.who.int/news-room/fact-sheets/detail/breast-cancer, accessed 9th Oct 2023

Research datasets

Three Phase III trials investigating Ribociclib combination with hormonal therapy vs. hormonal therapy alone in broad patient population with HR+, HER2- mBC

Over 2,000 patients enrolled between 2013 and 2016

Multimodal datasets anonymized in this collaboration included:

- Clinical data: demographics, diagnosis of cancer, medical conditions, lab assessments etc.
- Radiological data: CT/MRI scans with a subset containing lesion annotations from radiologists
- Genomics data: gene expression from archival tumor and ctDNA from plasma samples

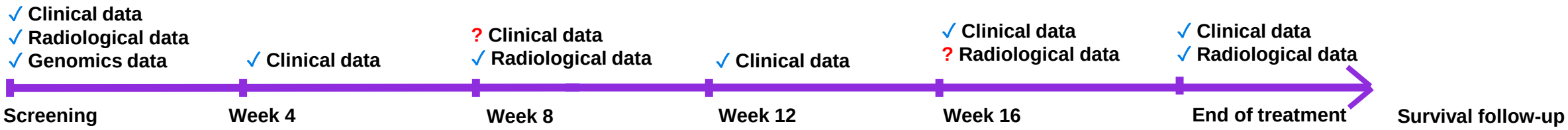


Figure: a possible scenario for data availability during patient journey in the trial.

Clinical data

Collected to address the efficacy and safety objective of clinical trials

Analysis Data Model (ADaM) with tabular structure

Industry standard meeting regulatory requirements

Low proportion of missingness, however data collected vary slightly across trials

Typically, one row per patient / assessment time / parameter

Patient ID	Age	Height	Weight
1000003	67	162	65
1000004	55	170	62
...	...	...	...

Patient ID	Visit	Parameter	Value (U/L)
1000003	Baseline	AST	25
...	Baseline	ALT	30
	Week 4 Day 1	AST	21
	Week 4 Day 1	ALT	29
	...	...	...

Radiological data

3D images in DICOM format, with a subset containing lesion annotations

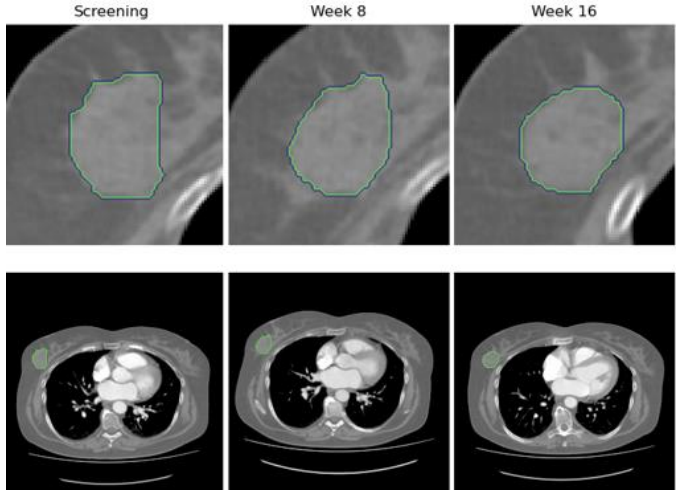
Most images are CT scans

A typical size of 512x512x~100 voxels (H, W, D)

Serves the basis for the detection of disease progression

Contain the raw information beyond lesion diameters

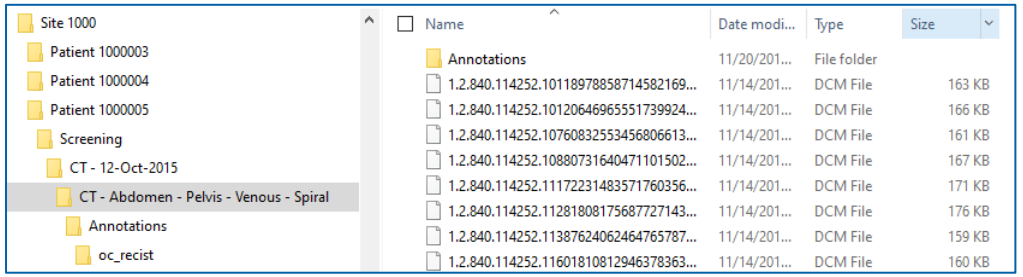
Features of interest not readily available



Require intensive data wrangling prior to analysis

Image quality varies depending on acquisition method, contrast

About 5TB of DICOM files



Genomics data

Archival tumor biopsy collected prior to treatment

Gene expression generated using Nanostring technology

Targets >700 genes and enriched for genes implicated in breast cancer and immuno-oncology

Plasma samples collected prior to treatment to assess ctDNA

NGS-based assay was used to sequence >500 genes from cancer related pathways

High-dimensional sparse data

In ctDNA data, many genes are infrequently mutated leading to sparsity in the data

For gene expression, genes can be highly correlated with one another

Require feature engineering and/or dimension reduction approaches prior to modeling

ADG... Inte...	AFF2 Integ...	AJUBA Integer	AKT1 I...	AKT2 Integ...	AKT3 Integ...	ALK Int...	ALOX1... Integer	AME... Integ...	ANAP... Integ...	...	TMB String	PURITY String
0	0	0	1	0	0	0	0	0	0		1.318659	0
0	0	0	1	0	0	0	0	0	0		0.6615	0.23
0	0	0	1	0	0	0	0	0	0		1.318728	0.09505
1	0	0	1	0	0	0	0	0	0		14.469734	0.18

Before modeling

- **Conduct exploratory data analysis**

Understand data available with respect to research question

- **Establish data usage**

Partitions train, validation and test prior to any modelling

- **Accommodate multimodal datasets**

Align on user-friendly analysis pipeline and computing platform

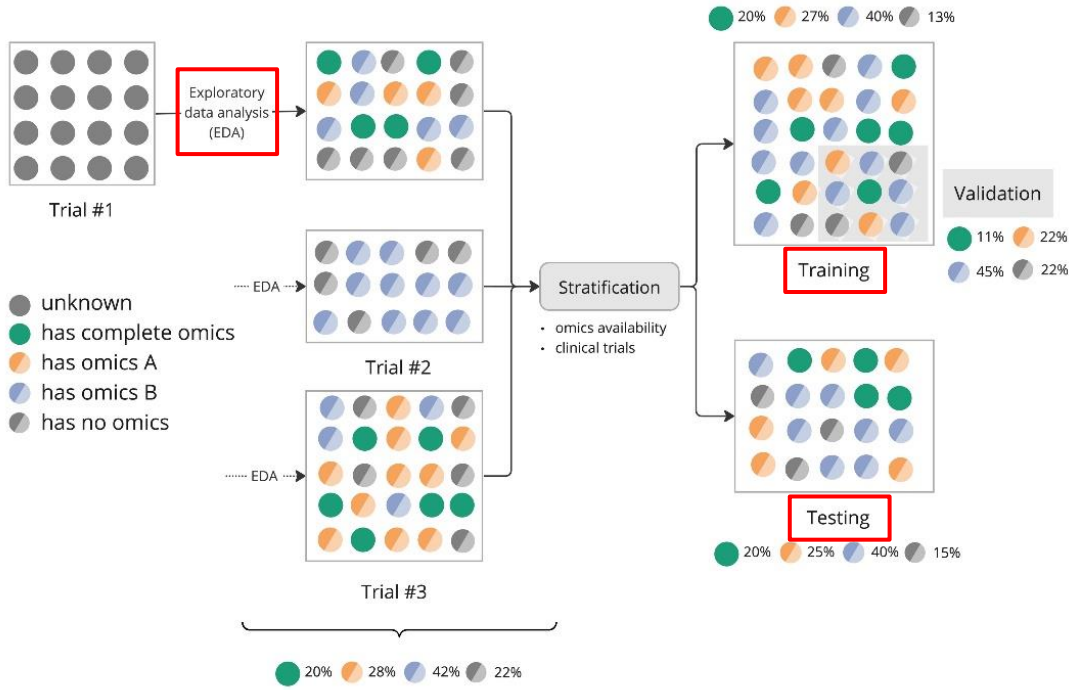


Figure: schematics of establishing data usage.

During & after modeling

Design rules that ensure a robust model training process and consistent comparisons between models

Flexible for uni/multi-modal training

Controls usage of testing dataset

Follow good data science practice

Write re-usable code

Share across language

Interepret models to make clinical impact

Imaging data implies using ML/AI algorithms

Model fusion further complicates interpretation

Engage with clinicians to understand the “target interpretation”

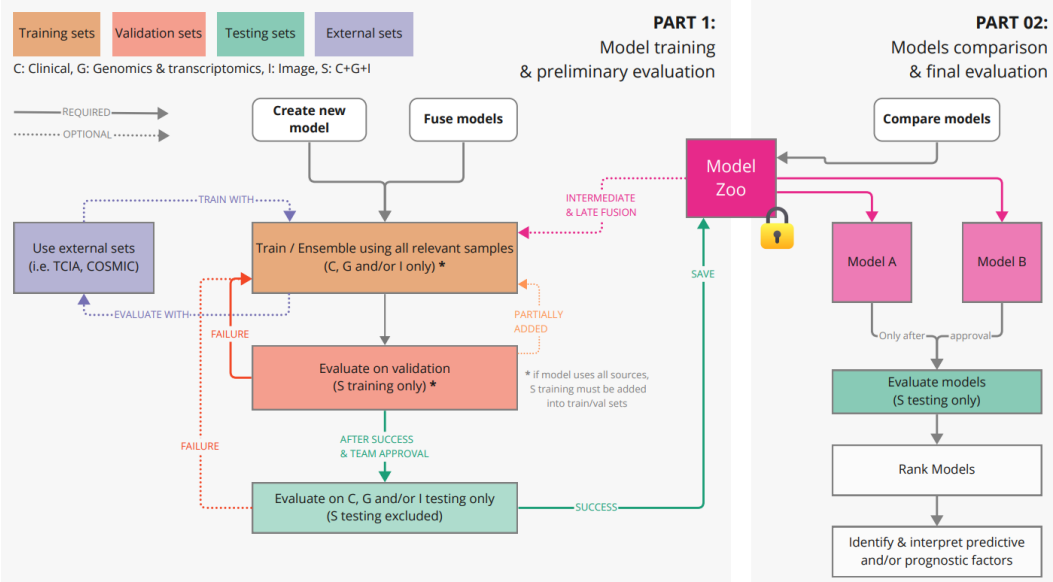


Figure: flow diagram to train, fuse and compare models.

Research outputs

***Multi-omics investigation on the prognostic and predictive factors in metastatic breast cancer using data from Phase III ribociclib clinical trials: A statistical and machine learning analysis plan*, published on medRXiv**

Establish a prospective analysis plan to wrangle, analyze, compare and combine models

***Multi-Center Collaboration with Regulated Industry to Pilot Approaches to Receiving and Analyzing Large Datasets*, presented at FDA Scientific Computing Days 2023**

Sharing and Collaboration in the Data Multiverse: Scientific Computing for Public Health Solutions

***Methodology for good machine learning with multi-omics data*, a tutorial paper accepted at Clinical Pharmacology & Therapeutics**

Help other scientists exploring complex omics data in a principled way

***A radiomics model for prediction of metastatic breast cancer progression risk*, abstract accepted at SPIE Medical Imaging 2024 Conference**

Demonstrate delta radiomics significantly improved PFS prediction over screening radiomics model

Take-home messages

Prior to analysis, conduct extensive data exploration

To assess data availability / quality for different research questions

Multimodal research require interdisciplinary team with domain experts

Extra efforts to ensure results are interpretable and clinically impactful

Thank you

NOVARTIS | Reimagining Medicine