

Data Review of RECIST Criteria using ML/AI

A Proof of Concept Example

Priya Gopal, TESARO, a GSK Company

Christine Teng, Merck & Co., Inc.

Sam Tomioka, Sunovion Pharmaceuticals

Introduction

-
- AI/ML is a growing area used to solve complex problems across different domains
 - Oncology is a growing field and lot of researchers are trying to find cure for different types of cancer
 - RECIST criteria is one of the important assessment that determines the efficacy of the medication for the particular type of tumor
 - The object of this POC, is to determine whether AI/ML technology could be used to review tumor measurements to determine the corresponding RECIST criteria for the clinical trial data

- Response Evaluation Criteria in Solid Tumors (RECIST) is a set of standardized rule to assess tumor responses after the start of the treatment
- The main response categories are:

Response		Definition
Complete Response	CR	All non-nodal target lesion (TL) s disappeared; all lymph nodes short axis <10mm
Partial Response	PR	Sum of diameters (SOD) decreased $\geq 30\%$ from baseline
Progressive Disease	PD	SOD increased $\geq 20\%$ from nadir and the 20% has absolute increase $\geq 5\text{mm}$
Stable Disease	SD	Not PR nor PD
Not Evaluable	NE	Cannot determine TL response

- There are 3 types of lesion in overall response assessment by RECIST :Target lesion, Non-target lesion and New lesion

Introduction

Target Lesion and New Lesion

- Target Lesion is measured at every visit.
- The main measurements include
 - Longest Diameter (LD) for non-nodal target lesion
 - Short Axis (SA) for target lymph node
 - Sum of Diameters (SOD) (LD + SA)
 - Percentage of change from nadir SOD
 - Percentage of change from baseline SOD



“nadir” is the smallest SOD prior to current assessment.

Lesion identified post-baseline is new lesion and is considered PD.

This POC, does not consider non-target lesion for determining the RECIST criteria.

Introduction

Site vs Central Review

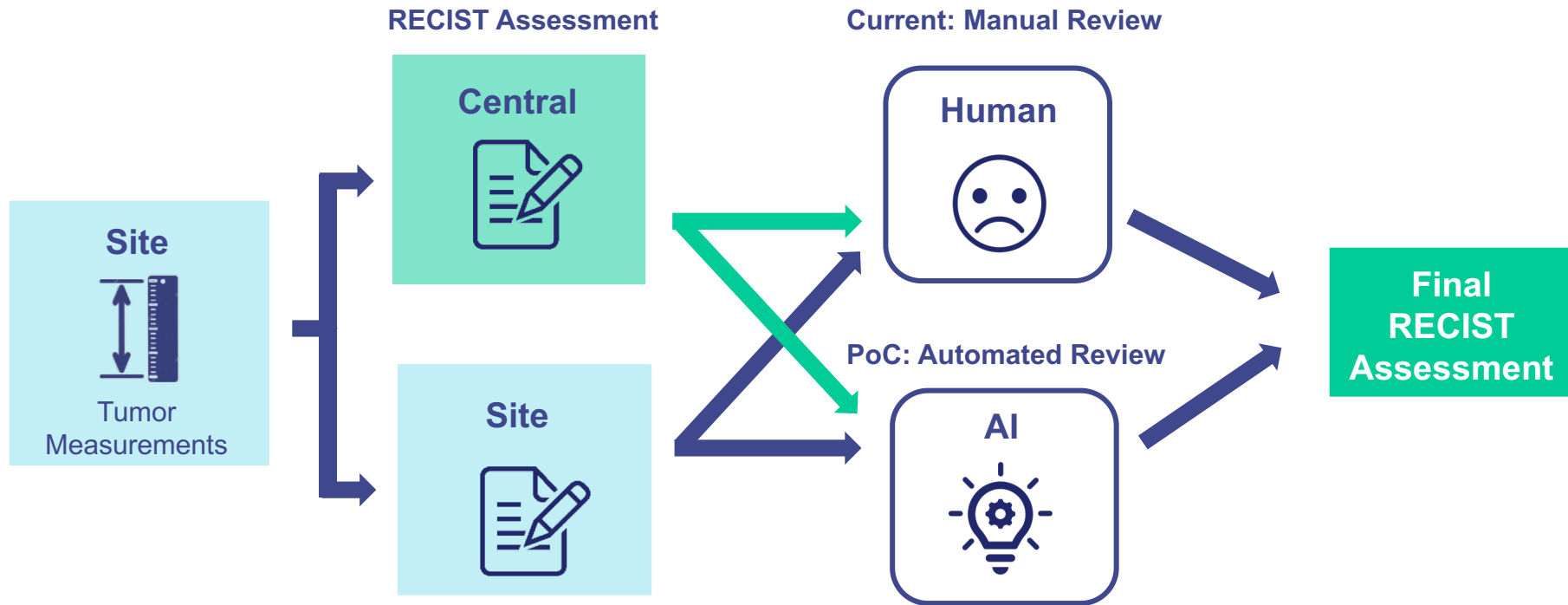
- Site data is collected on the CRF
- Central Review is done by group of independent trained readers.
- There are usually differences between site and central evaluation
- The differences are at both visit-level and in overall assessment evaluation
- Data management and medical writers need to understand the causes of differences to determine the correct assessment for each subject



AI/ML

Supporting heading

Motivation



Introduction

- Machine Learning is an application of Artificial Intelligence (AI) that provides ability for the machines to learn from experience and data patterns without explicit programming
- There are 3 major categories of Machine Learning (ML):
 - **Supervised learning algorithms** are used when the data contains input and set of desired output (labeled data). Data classification algorithms and regression algorithms are part of supervised learning model.
 - **Unsupervised learning algorithms** are used when the machine is trained using unlabeled data. The data is grouped or clustered into different categories to find structure in the data.
 - **Reinforcement learning algorithms** is different when compared to supervised and unsupervised learning. It continuously learns from the environment in an iterative fashion.
- For this POC, we are using supervised ML algorithms to predict the RECIST criteria.

Response Criteria Prediction

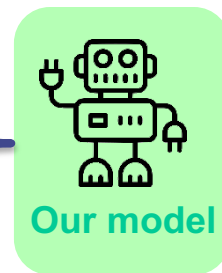
Supporting heading

Response Criteria Prediction

Training Data structure

Input

Features	
BSUM	Baseline Sum of Diameter
SUMDIAM	Sum of Diameter (Sum of Longest Diameters and Sum of diameter of non-lymph-node tumor) at current visit
PCBSD	Percent Change From Baseline in Sum of Diameter
NADIR	Smallest Sum of Diameter prior to current time point
ACNSD	Absolute Change From Nadir in Sum of Diameter
PCNSD	Percent Change From Nadir in Sum of Diameter
NEWLSN	New Lesion (1=yes,0=no)



Output

Target Lesion Responses

CR

PR

PD

SD

NE

Imputation

Imputing Missing Data

	If missing	Assumption
NADIR	BSUM	No post baseline measurement taken
PCBSD	$100 * (\text{SUMDIAM} - \text{BSUM}) / \text{BSUM}$	
SUMDIAM	Dynamically chosen based on minimum loss of the model	Label is NE
NEWLSN	0 (no new lesion)	

Additional imputation methods were explored to test the generalization of the above rules.

Preparing the Data

Site and Central Data

- Site (n=695) and central (n=920) data are used for evaluating different ML algorithms
- The datasets were prepared as follows.

	Training and Validation Data	Test Data	Test Data ID
1	Central	site	A
2	Site	central	B
3	central85%+site85%	central*15%	C
4	central85%+site85%	site*15%	D

Preparing the Data

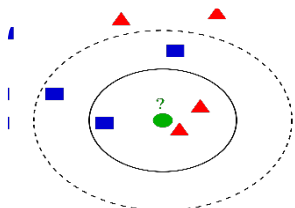
Central Data

- Central Data is usually used to determine primary endpoint. The model used to train using Central Data can identify inconsistency in response assessment more accurately, than site data.
- The models were prepared using 70% of the Central Data, and tested with either site data, or central 30%.

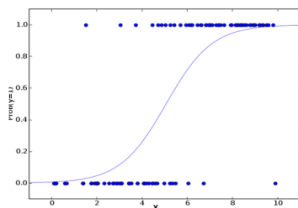
	Training and Validation Data	Test Data	Test Data ID
5	central*70%	central*30%	E
6	central*70%	site	F

ML Algorithms

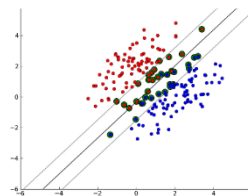
Initially these 5 classifiers are explored in this PoC.



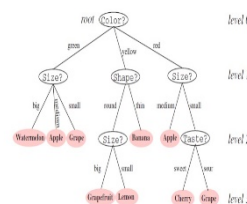
**K Nearest
Neighbour**



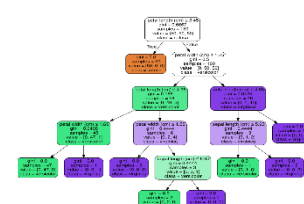
**Logistic
Regression**



**Support
Vector
Machines**



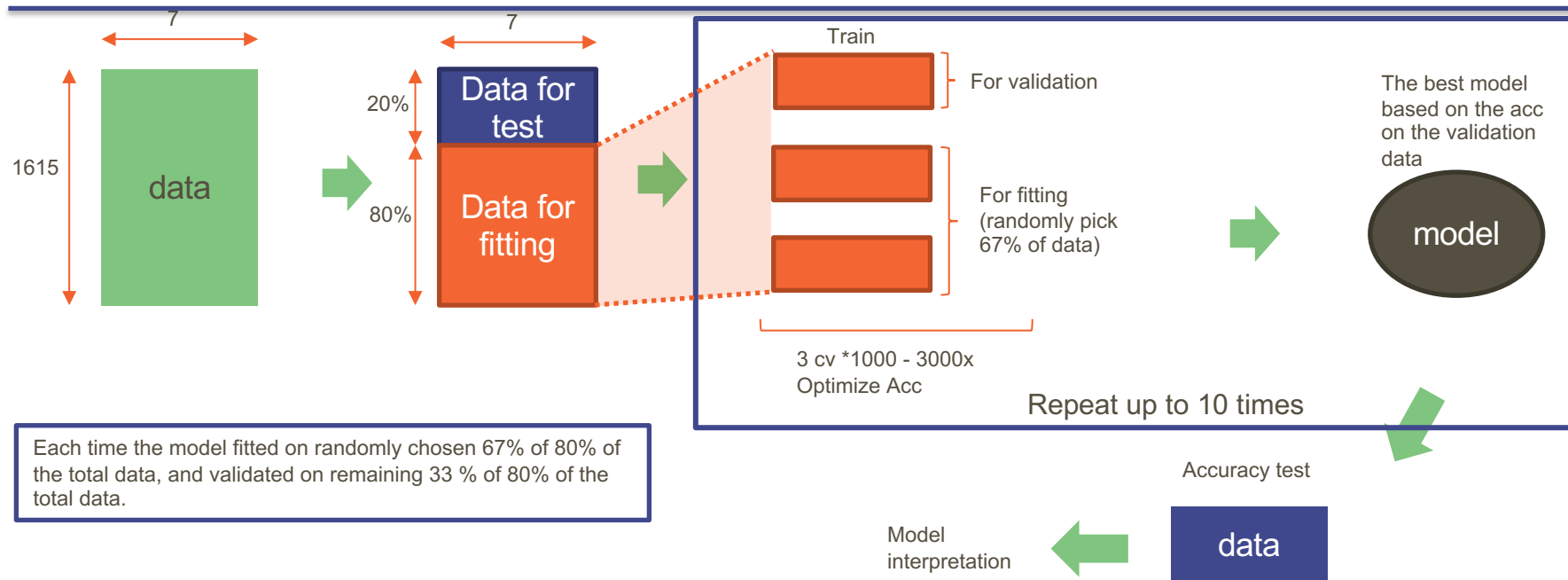
**Random
Forest**



**Gradient
Boosting**

ML Algorithms

Model training and testing data configurations



Results and Comparison

- The results and comparison of testing and validating the data using different ML models are as follows:
- The highest level of accuracy is using random forest or Xgboost algorithm using Test data described in D

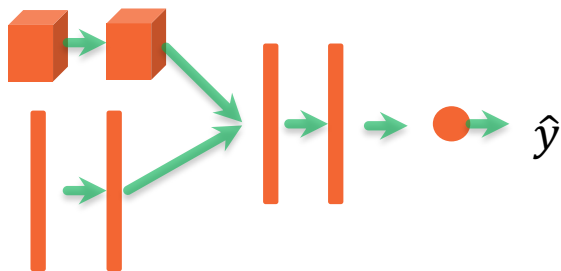
	Test Data ID	Test Data	metric	rf	svc	lr	knn	Xgb
1	A	site	acc	90.1	84.5	88.1	84.6	90.4
2	B	central	acc	83.0	73.9	78.0	81.1	82.3
3	C	central*15%	acc	82.6	73.9	81.2	82.6	84.1
4	D	site*15%	acc	95.2	86.5	90.3	93.3	95.2
5	E	central*30%	acc	84.4	72.1	81.5	85.9	85.9
6	F	site	acc	88.4	82.6	87.9	87.4	89.2

Neural Networks

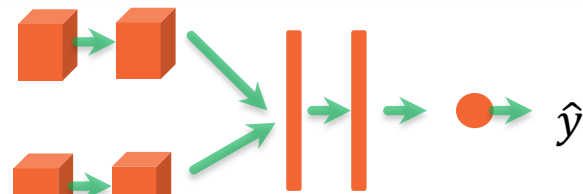
Neural Networks architectures to estimate RECIST outcome on target lesion response



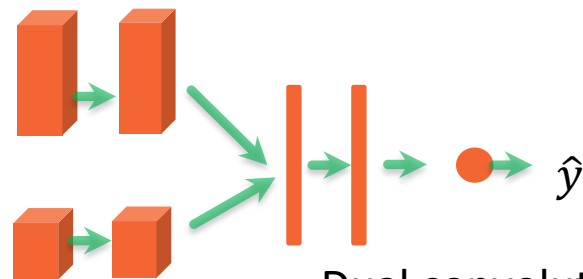
Arbour et.al. reported at ASCO the 4-layer neural network architecture (ml-RECIST) can estimate RECIST on overall response



Convolution + feed forward



Parallel convolution



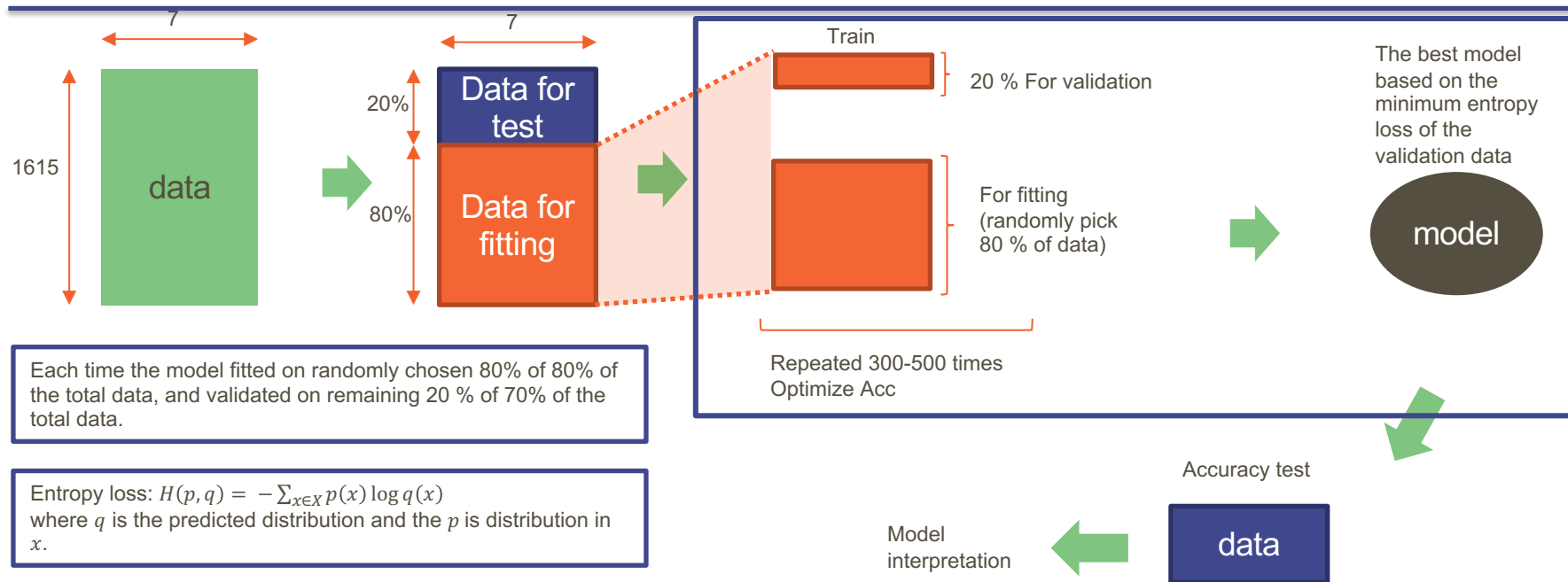
Dual convolution

[1] C. Arbour, Kathryn et. al., *ml-RECIST: Machine learning to estimate RECIST in patients with NSCLC treated with PD-(L)1 blockade*. Journal of Clinical Oncology 2019 37:15_suppl, 9052-9052



Neural Networks

Model training and testing data configurations



Neural Networks

Results and Comparison

- The **ml-RECIST** performed well in D.
- **Parallel convolution** achieved the best performance

	Neural network	Test Data ID		Acc	XGB
7	ml-RECIST[1]	C	central*15%	86.2	84.1
8	ml-RECIST	D	site*15%	97.1	95.2
9	fnn+ convolution	C	central*15%	87.0	84.1
10	fnn+ convolution	D	site*15%	97.1	95.2
11	Parallel Convolution	C	central*15%	86.2	84.1
12	Parallel Convolution	D	site*15%	98.0	95.2
13	Dual Convolution	E	central*30%	85.8	85.9
14	Dual Convolution	F	site	92.0	89.2

[1] C. Arbour, Kathryn et. al., *ml-RECIST: Machine learning to estimate RECIST in patients with NSCLC treated with PD-(L)1 blockade*. Journal of Clinical Oncology 2019 37:15_suppl, 9052-9052

-
- Parallel convolution neural network trained on 85% of central and +85% of site estimated RECIST assessment with accuracy of **98.0%** of unseen site data.
 - In general, the neural network performed better than the traditional classification algorithms.
 - This approach may be used to determine the RECIST outcomes consistently and efficiently.

Question?

Thank You