# A late fusion machine learning approach utilizing TruTumor ex-vivo multi-level histopathology data for chemotherapy response prediction in solid tumors

Pradeepa S[1*], Koushika R[1], Poojitha N[1], Dr. Vijay Pillai[2], Dr. Yogesh Dokhe[2], Dr. Vivek Shetty[2], Dr. Vidya Bhushan[2], Dr. Rohit Ranade[2], Dr. Akshita Singh[2], Oliyarasi M[1], Ritu Malhotra[3], Dr. Lalitha Laxhmi[2], Govindaraj K[1], Pavithira [1], Mohit Malhotra[3], Satish Sankaran[1#]

1. Farcast Biosciences, Bangalore, Karnataka, India, 2. Narayana Health, Bangalore, India, 3. Farcast Biosciences, Pensacola, FL

* Presenting Author , # Corresponding author: Satish.Sankaran@farcastbio.com

## Background

Reliable treatment response prediction is essential for guiding chemotherapy decisions in solid tumors. Patient-derived tumor explant models such as the Farcast TruTumor platform are promising tools for assessing individual drug responses. However, small cohort sizes with high tumor heterogeneity leads to overfitting and limited generalization in single-level models, failing to capture variance contributed by multiple histopathological features. There is a need for modeling strategies that can integrate these diverse information while remaining robust under low-label clinical settings.

## Study Strategy

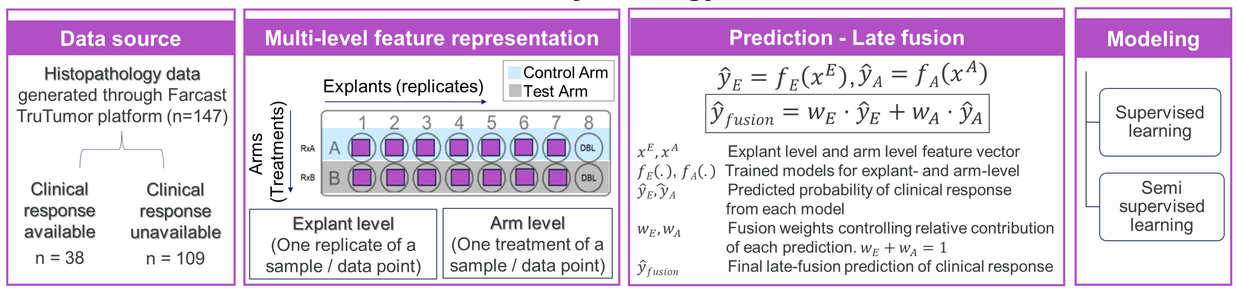



**Figure 1**: Study strategy implemented for heterogeneity-aware, multi-level chemotherapy response modeling using histopathology assay outputs from the Farcast TruTumor platform under a low-label clinical setting.

## Methods

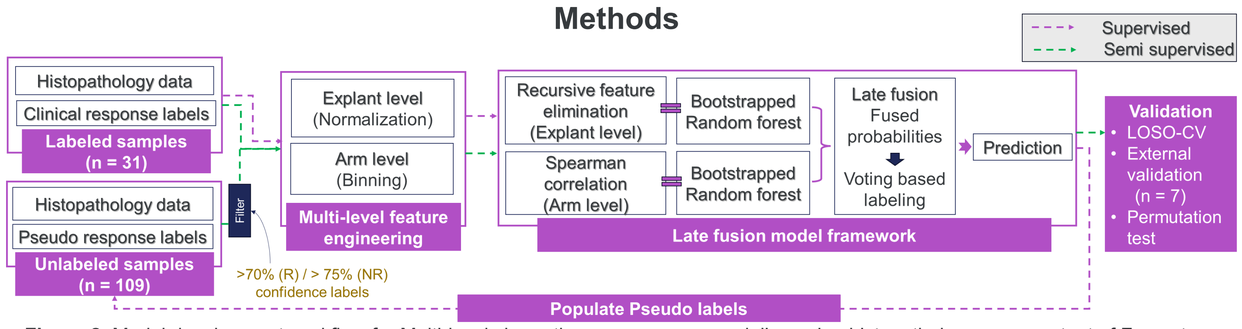



**Figure 2**: Model development workflow for Multi-level chemotherapy response modeling using histopathology assay output of Farcast TruTumor platform

### Data generation

Fresh, surgically resected Head and Neck (HNSCC), Ovarian cancer (CaOv), Breast cancer (CaBr) samples were collected from consented patients along with matched blood samples. The sample was processed to generate thin explants. Tumor explants were cultured with media and autologous plasma for 72 hours, with chemotherapy treatment every 24 hours.

### Multi-Level Feature Engineering

Histopathological features were organized at two levels:

- Explant level, capturing treated versus control differences within samples, and
- Arm level, summarizing treatment-group effects using statistical, fold change, variance, and PCA based features.

### Model Training and Feature Selection

Bootstrapped Random Forest models were trained independently at each level. Feature selection used,

- Recursive feature elimination for explant level features and
- Spearman correlation filtering for arm level features.

### Semi Supervised Learning Strategy

Semi supervised learning was implemented via pseudo-labeling of unlabeled explants. Calibrated prediction probabilities were filtered to retain high confidence samples (probability > 0.7 for Responder (R) and probability > 0.75 for Non-Responder (NR)) for model retraining.

### Late Fusion and Evaluation

Final response predictions were obtained using a late-fusion strategy, combining calibrated probabilities from explant and arm level models through weighted averaging. Model performance was evaluated using patient-level leave-one-sample-out cross-validation (LOSO-CV) on the labeled cohort, ensuring independence between training and test samples.

### Generalization tests

- External validation: To assess model robustness and generalization capability, predictions were evaluated on an independent external validation cohort. Performance metrics were computed for both models. Sample level prediction outcomes were compared using net gain in prediction histograms.
- Permutation tests: Statistical significance of observed gains was assessed via permutation testing, iterations = 50.

## Results

### Exploratory data analysis of clinically labeled samples (n=38)

A

| Parameter | Category | | Value (No. of samples) | Value (% of samples) |
|---|---|---|---|---|
| Cancer Type | HNSCC | Cisplatin | 24 | 68.4 |
| | | Carboplatin + Paclitaxel | 1 | |
| | | Carboplatin | 1 | |
| | CaOv | Carboplatin + Paclitaxel | 10 | 26.3 |
| | CaBr | Carboplatin + Paclitaxel | 2 | 5.3 |
| Gender | Male | | 19 | 50 |
| | Female | | 19 | 50 |
| Age (years) | Median (Range) | | 46 (27–67) | |
| Clinical Response | Responder | | 26 | 68.4 |
| | Non Responder | | 12 | 31.6 |

B

| Feature | Description |
|---|---|
| Tumor (%) | Percentage of tissue area occupied by tumor cells |
| Necrosis | Tissue regions showing cell death due to damage |
| Pyknosis | Condensed cell nuclei indicating apoptosis |
| Discohesion | Loss of structural integrity between tumor cells |
| Caspase-3 (Cas-3) (%) | Percentage of tumor area with the marker of programmed cell death (apoptosis) |
| Immune component (%) | Area occupied by the immune cells in the tissue as a percentage |
| Tumor infiltrated immune cells (%) | Average area occupied by the immune cells near the tumor border as a percentage |

C D

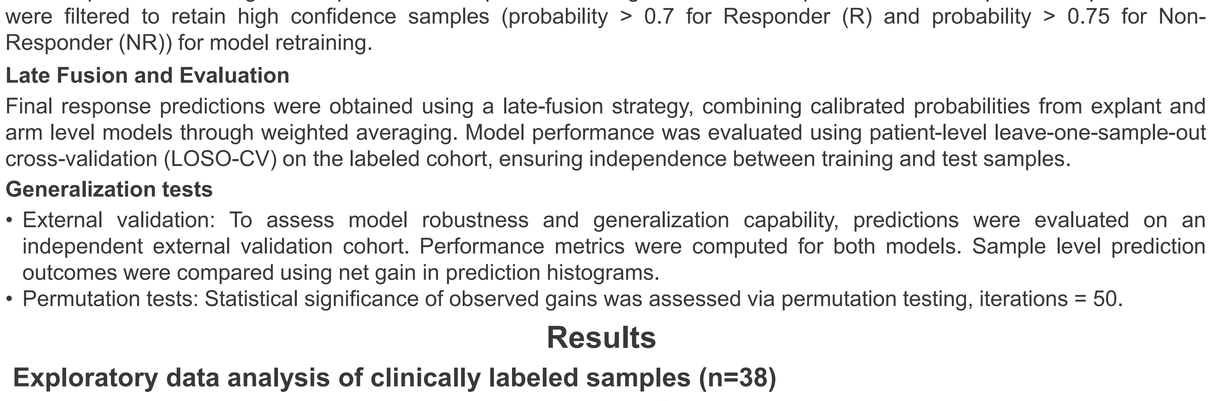



**Figure 3**: **Biological signal and response heterogeneity across explant level features**. (A) Metadata based stratification of clinically labeled samples. (B) Description of histopathological features used for modeling. (C) Correlation matrix of explant level features used for redundancy removal and feature selection. (D) PCA projection reveals distinct separation of responder and non responder samples in arm level compared to explant level, supporting heterogeneity-aware modeling.

### Selected Histopathological Features for Model Development

A

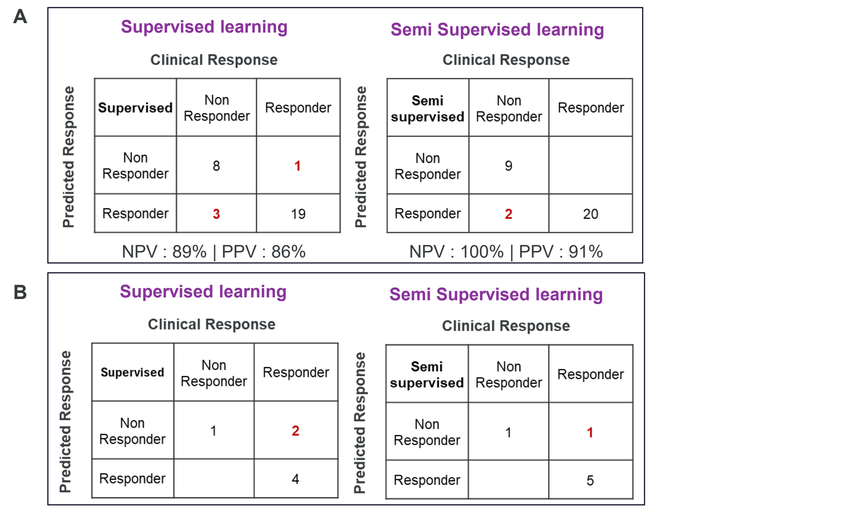
| Feature Level | Feature | Feature definition and Computation | Expected Trend in Responders |
|---|---|---|---|
| Arm-level | FC (Cas-3 %) | $Score\ (FC) = \begin{cases} +1, & FC \geq 1.2 \\ +0.5, & 1.0 \leq FC < 1.2 \\ -0.5, & 0.8 < FC < 1.0 \\ -1, & FC \leq 0.8 \end{cases}$ | ↑ Increase |
| | FC (Tumor %) | | ↓ Decrease |
| | Weighted Pyknosis | $Weighted\ (x) = \log_2 FC(x)(1-p_x)$ where, $p_x$ : Mann–Whitney U ($treated_x$ , $control_x$) | ↑ Increase |
| | Weighted Discohesion | | ↑ Increase |
| | Weighted Cas-3 | | ↑ Increase |
| | SDdiff (Tumor %) | $SDdiff\ (x) = SD(treated_x) / SD(control_x)$ | ↓ Decrease |

B

| Feature Level | Feature | Feature definition and Computation | Expected Trend in Responders |
|---|---|---|---|
| Explant level | Tumor (%) | Percentage (%) of tumor area | ↓ Decrease |
| | Necrosis | Scores | ↑ Increase |
| | Pyknosis | Scores | ↑ Increase |
| | Discohesion | Scores | ↑ Increase |
| | Cas-3 (%) | Percentage (%) of Cas-3–positive tumor cells | ↑ Increase |

**Figure 4: Selection of histopathological features used for model development at (A) arm level and (B) explant level.** Features were selected using recursive feature elimination and Spearman correlation-based filtering at the explant and arm level. The table summarizes feature definitions, calculation methods, and the expected directional trend in clinical responders. Abbreviations - *FC: fold change; SDdiff: standard deviation ratio.*

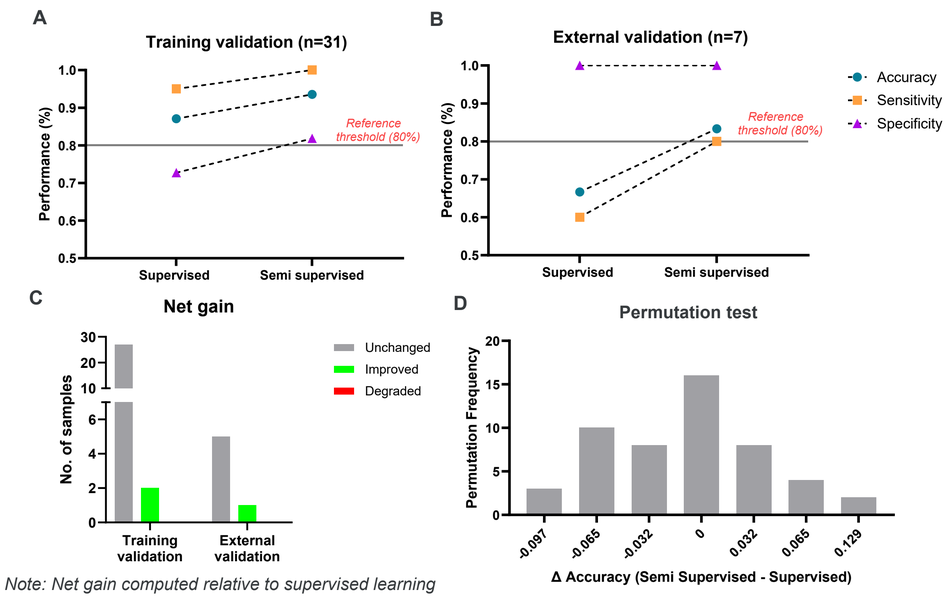
## Model Performance : Semi Supervised vs Supervised Learning




*Note: NPV and PPV not shown for external validation due to limited sample size.*

**Figure 5**: Confusion matrices show prediction performance during internal training validation, n=31 (A) and independent external validation, n=7 (B). Semi supervised learning demonstrates improved negative and positive predictive values compared to supervised learning across validation settings. *Note – Numbers inside matrices indicate sample counts*

## Consistent Performance Gains with Semi Supervised Learning




*Note: Net gain computed relative to supervised learning*

**Figure 6**: **Performance comparison of supervised and semi supervised learning frameworks.** (A–B) Sensitivity, specificity, and overall accuracy across internal training and independent external validation cohorts. (C) Net gain analysis illustrating sample-wise improvement in prediction accuracy under semi supervised learning relative to supervised models. (D) Permutation testing (50 iterations) shows that under random label assignments, the semi supervised model does not outperform the supervised model, indicating no artificial performance gains.

## Semi Supervised Modeling Preserves Heterogeneous Response Patterns

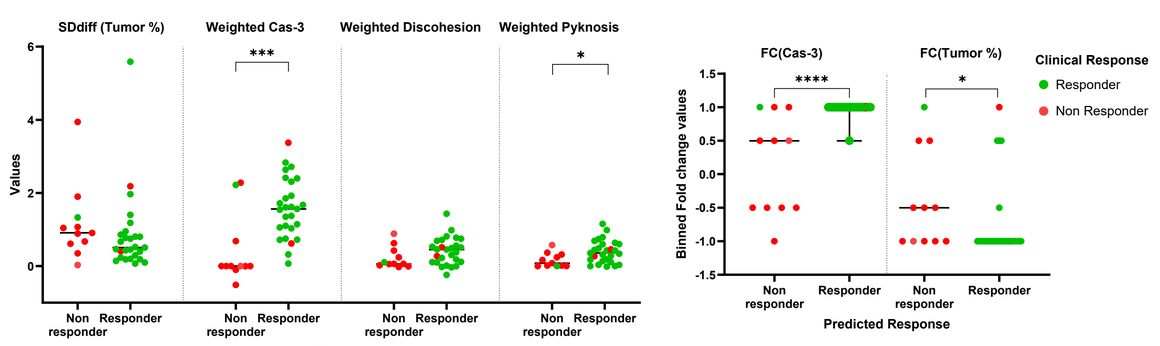



**Figure 7. Semi supervised modeling preserves heterogeneous response patterns across cohorts.** Predicted response distributions across samples (n = 38), stratified by clinical response. Distinct response-associated patterns are consistently retained under semi supervised modeling, indicating preservation of biologically relevant signal. Statistical significance was assessed using the Mann–Whitney U test.

## Semi Supervised Modeling Improves Prediction Beyond Manual Log2FC-Based Rules

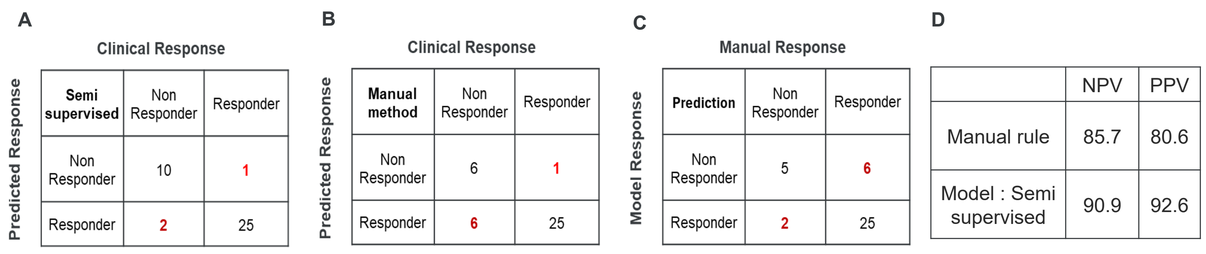



**Figure 8. Comparison of semi supervised modeling and manual Log2FC-based response prediction (n = 38).** (A–B) Confusion matrices comparing predicted versus clinical response for the semi supervised model and a manual Log2FC-based rule. (C) Prediction disagreement matrix highlighting samples, where the two approaches yield discordant predictions. (D) Comparison of negative and positive predictive values (NPV, PPV) demonstrates improved predictive performance under semi supervised modeling relative to the manual rule.

## Conclusion

Semi supervised learning enables robust prediction of clinical response by capturing heterogeneous response patterns within and across samples. By utilizing unlabeled data and maintaining feature-driven heterogeneity, the framework improves response stratification accuracy. These results support its translational potential for personalized medicine and improved patient outcomes.

## References

1. Basak et al (2024) Tumor histoculture captures the dynamic interactions between tumor and immune components in response to anti-PD1 in head and neck cancer. *Nat Commun* 15, 1585
2. Hédou et al (2024) Discovery of sparse, reliable omic biomarkers with Stabl. *Nat Biotechnol* 42, 1581
3. Lipkova, Jana, et al (2022) Artificial intelligence for multimodal data integration in oncology. *Cancer cell* 40.10, 1095-1110.
4. Zhu, X., & Goldberg, A. (2009). Introduction to semi-supervised learning. *Morgan & Claypool Publishers*.