

Supplement to “Personalised Neoadjuvant Therapy Recommendations for Breast Cancer: A Causal Inference Approach using Multi-omics Data”

1 Multi-omics Data

Table 1 shows a list of example feature classes present in the multi-omics data of breast cancer patients [Sammur et al., 2022].

Table 1: Some Features of Breast Cancer Patients

Feature Class	Attribute	Description
Clinical	Tumour size	Size of cancer tumour
Clinical	Age	Age of the patient at diagnosis
Clinical	Histological sub-type	Histological subtype of the tumour
Clinical	HER2 status	Human Epidermal Growth Factor Receptor status
Clinical	ER status	Estrogen Receptor status
DNA	TMB	Tumour Mutational Burden
DNA	HRD score	Homologous Recombination Deficiency score
RNA	PGR expression	Progesterone Receptor expression
RNA	ESR1 expression	Estrogen Receptor expression
RNA	Taxane score	Score indicating Taxane sensitivity
Digital pathology	Lymphocyte density	Number of Tumour-Infiltrating Lymphocytes (TILs) in the tumour
Treatment	Anti-HER2	Use of anti-HER2 drugs
Treatment	Anthracycline	Use of anthracycline drugs

Multi-omics data, including clinical features, digital pathology, genomic and transcriptomic profiles (DNA and RNA sequencing), were utilised to develop a machine learning model for breast cancer treatment. This model integrates multi-omics data to improve predictive capabilities by providing a comprehensive understanding of tumour molecular and cellular landscapes, surpassing traditional risk stratification methods [Hasin et al., 2017]. The study gathered 74 variables from breast cancer patients, encompassing clinical features, tumour genomic landscapes (DNA, RNA), digital pathology, and treatment details [Sammur et al., 2022].

Key clinical features include patient age, tumour size, histology, hormone receptor (ER and HER2) status, lymph node (LN) status, and ER Allred score. Digital pathology, represented by the number of Tumour-Infiltrating Lymphocytes (TILs) in the tumour (Lymphocyte density), offers insights into immune micro-environments [Sammur et al., 2022]. DNA-related features such as Tumour Mutational Burden (TMB), Homologous Recombination Deficiency score (HRD score), and key gene mutations (PIK3CA, TP53) reflect tumour heterogeneity and guide personalised therapies. RNA-related features include gene expression profiles of hormone receptors, HER2, immune responses, and tumour stemness, which influence treatment strategies and response prediction.

Treatment-related variables detail chemotherapy regimens, including the number of cycles and the specific drugs used in each cycle (taxanes, anthracyclines, and anti-HER2). The variables "any Anthracycline" and "any anti-HER2" indicate whether the patient received anthracycline or anti-HER2 therapy, respectively. These variables help assess the impact of different chemotherapy regimens on treatment response, providing valuable insights into the effectiveness of neoadjuvant therapy plans for breast cancer patients.

2 Algorithm for CTR

Following the CTR framework, we have devised the Causality-based Therapy Plan Recommendation (CTR) algorithm for personalised therapy plan recommendation, as depicted in Algorithm 1.

Algorithm 1 Causality-based Therapy Plan Recommendation (CTR)

— Training Causal Tree Models —

Input: Training dataset D with a set of attributes X , therapy plans T , and outcome Y .

Output: Causal Trees set $Z = \{Z_1, Z_2, \dots, Z_q\}$.

Let $Z = \emptyset$

for each therapy plan $TP \in T$ **do**

Fit a Generalized Linear Model (GLM) to estimate propensity scores.

Call the Causal Tree method with TP as the treatment variable, outcome Y , and X as the covariate set.

Store the obtained causal tree in Z as Z_k .

end for

return Causal Trees set Z for each therapy plan.

— Personalized Recommendation —

Input: Dataset D_R with a set of attributes A , Causal Trees set $Z = \{Z_k\}$.

Output: Recommendation set $R(A, TP, CATE_{TP})$, where TP is the therapy plan, and $CATE_{TP}$ is the estimated conditional average treatment effect.

Let $R = \emptyset$

for each individual $i \in D_R$ **do**

Retrieve attribute values $A(i)$.

for each causal tree $Z_j \in Z$ **do**

Retrieve causal factor TP_k of causal tree Z_k .

if value of attribute $TP_k(i) = 0$ **then**

Search for the subset that individual i belongs to.

Retrieve $CATE_i(TP_k)$.

end if

end for

Select the therapy plan TP with the largest $CATE_i(TP)$ to recommend for individual i .

Add tuple $(A(i), TP, CATE_i(TP))$ into R .

end for

return R .

3 Recovery Rate in each Breast Cancer Subtype

Breast cancer is categorised into five main subtypes by the PAM50 subtyping technique: Luminal A, Luminal B, HER2-enriched, Basal-like (Triple-Negative), and Normal-like. In this study, we conducted a Recovery Rate evaluation for each breast cancer subtype to assess the effectiveness of the proposed CTR method. The experimental results, summarised in Figure 1, demonstrate that the proposed CTR method outperforms both the Current Protocol and the best-performing baseline method, Random Forest, particularly in the most challenging subtype, Triple-Negative. These findings highlight the CTR method’s capability to provide personalised neoadjuvant therapy recommendations that improve treatment outcomes for patients across all breast cancer subtypes.

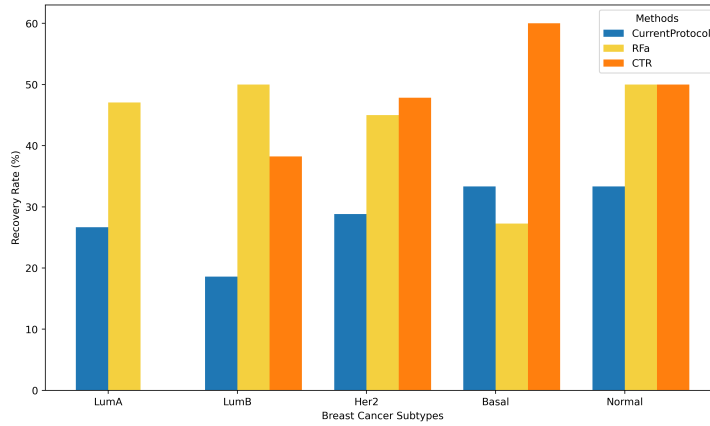


Figure 1: Comparison of Recovery Rates in each Breast Cancer Subtypes

4 The application of the CTR model for an alternative setting

4.1 Neoadjuvant Therapy Sequence

In practice, Taxane and Anthracycline are administered as a block-sequential regimen. After selecting the number of drugs to be used, the next step is to determine the therapy sequence for patients receiving both Taxane and Anthracycline. Specifically, we need to decide which drug should be administered first and which second. Our CTR model can provide recommendations for the therapy sequence using a similar framework. We denote three combinations of block-sequential Taxane and Anthracycline as follows:

- Therapy Sequence 1 (TS1): Taxane is administered first ($T = 1$) and Anthracycline second ($T = 0$), where $A = 1$ indicates that Anthracycline is administered in any cycle.
- Therapy Sequence 2 (TS2): Anthracycline is administered first ($T = 1$) and Taxane second ($T = 0$).

- Therapy Sequence 3 (TS3): Only Taxane is used, with no Anthracycline administered ($A = 0$).

In this case, TS3 is identical to Therapy Plan 3 in the previous experiment setting which use only Taxane, allowing us to use it for cross-verification. For a breast cancer patient receiving recommendations from both models, TP3 and TS3, it provides additional confidence in applying only Taxane as the therapy plan.

We also compared this experimental setting to the current protocol. The results in Figure 2 show that our model outperforms the current protocol in terms of Recovery Rate, and Figure 3 demonstrates superior AUUC performance compared to all baseline methods.

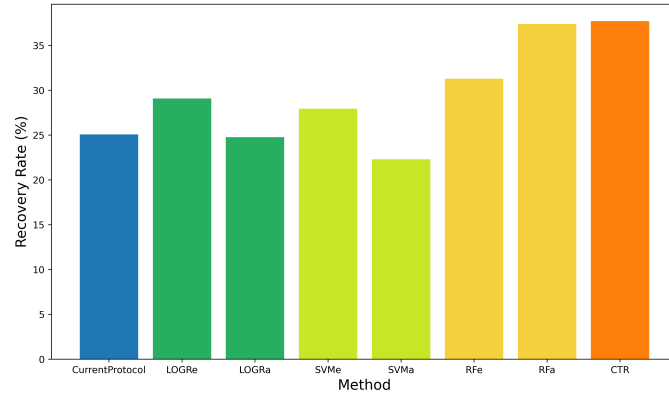


Figure 2: Comparison of Recovery Rates for Therapy Sequence setting

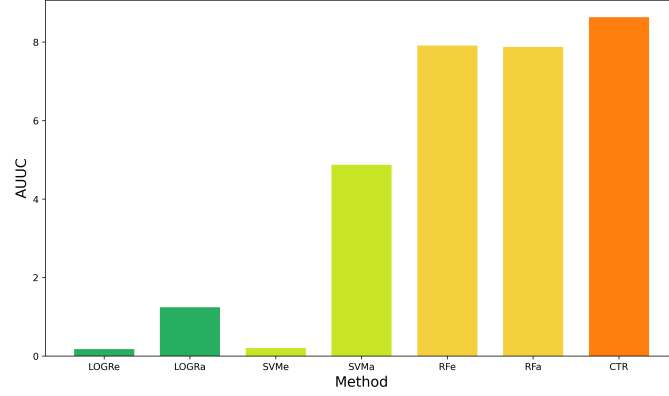


Figure 3: Comparison of Model Performance for Therapy Sequence Setting

4.2 Web App for Neoadjuvant therapy Sequence Recommendation

The Web App for Neoadjuvant Therapy Sequence Recommendation is deployed online, providing another option for neoadjuvant therapy recommendation ¹.

Neoadjuvant Therapy Sequence (TS) Recommendation App

Upload data for Breast Cancer Patient features (CSV File)

Browse...

new-BreastCancer-patients.csv

Upload complete

Recommend

Download Recommendation (session/007930ac2a7041df02fd19b9de59dff1/)

Recommendation Results

Show 5 entries

Search:

	Patient.ID	Recommendation
1	T198	Optimal TP: Anthracycline first, Taxane second Alternative TP: Only Taxane
2	T208	Optimal TP: Taxane first, Anthracycline second Alternative TP: Only Taxane
3	T211	Optimal TP: Anthracycline first, Taxane second Alternative TP: Only Taxane
4	A005	Optimal TP: Taxane first, Anthracycline second Alternative TP: Only Taxane
5	A010	Optimal TP: Anthracycline first, Taxane second Alternative TP: Taxane first, Anthracycline second

Showing 1 to 5 of 10 entries

Previous

1

2

Next

Figure 4: Web App for Neoadjuvant Therapy Sequence Recommendation

The Figure 4 show the web apps for and Neoadjuvant Therapy Sequence Recommendation, providing consistent results. For example:

- The patient T198 is recommended to apply 3 drugs: Taxane, Anthracycline, and anti-HER2, with the sequence Anthracycline administered first and Taxane second.
- The patient A208 was recommended to apply two drugs: Taxane and Anthracycline, with the sequence Taxane administered first and Anthracycline second.

¹<https://gc8kwh-tuyen-vu.shinyapps.io/causalR-TS/>

References

- Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome Biology*, 18:1–15, 2017.
- S.-J. Sammut, M. Crispin-Ortuzar, S.-F. Chin, E. Provenzano, H. A. Bardwell, W. Ma, W. Cope, A. Dariush, S.-J. Dawson, J. E. Abraham, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*, 601(7894): 623–629, 2022.