Conformal Bidirectional Floating Search (CBFS): Pseudocode

Algorithm 1: CBFS (main procedure)

Input: Feature indices $\{1, \ldots, d\}$; target subset size k; Max Iterations $I_{\max}$; Max Patience $P_{\max}$.

Output: Best subset S_{best} and best score $m_{\text{best}} = \mathcal{M}(S_{\text{best}})$.

$S \leftarrow \emptyset$; $U \leftarrow \{1, \ldots, d\}$; $U^* \leftarrow \emptyset$;

Until $|S| = k$ **do:**;

  $(S, U, U^*) \leftarrow$ BidirectionalSearch(S, U, U^*, k);

  $U \leftarrow (U \cup U^*) \setminus S$; $U^* \leftarrow \emptyset$;

$m \leftarrow \mathcal{M}(S)$; $S_{\text{best}} \leftarrow S$; $m_{\text{best}} \leftarrow m$;

$patience \leftarrow 0$; $iter \leftarrow 0$;

while $iter < I_{\max}$ ***and*** $patience < P_{\max}$ **do**

  $iter \leftarrow iter + 1$;

  $(S, U, m, S_{\text{best}}, m_{\text{best}}, patience) \leftarrow$ RefineStep$(S, U, m, S_{\text{best}}, m_{\text{best}}, patience)$;

return $S_{\text{best}}, m_{\text{best}}$;

Algorithm 2: BidirectionalSearch

Input: Current S, pools U and U^*, target size k.

Output: Updated (S, U, U^*) with $|S| = k$.

while $|S| < k$ **do**

  $R_{add} \leftarrow$ RankJMI(U, S); $f_{add} \leftarrow \text{top}(R_{add})$;

  $S \leftarrow S \cup \{f_{add}\}$; $U \leftarrow U \setminus \{f_{add}\}$;

  if $U \neq \emptyset$ **then**

    $R_{remove} \leftarrow$ RankCRFE(U); $f_{prune} \leftarrow \text{top}(R_{remove})$;

    $U \leftarrow U \setminus \{f_{prune}\}$; $U^* \leftarrow U^* \cup \{f_{prune}\}$;

return (S, U, U^*);

Algorithm 3: RefineStep (floating adjustment, one iteration)

Input: S, U, current score $m = \mathcal{M}(S)$; global best $(S_{\text{best}}, m_{\text{best}})$; *patience*.
Output: Updated $(S, U, m, S_{\text{best}}, m_{\text{best}}, \textit{patience})$.

$R_{remove} \leftarrow \text{RankCRFE}(S)$; $f_{remove} \leftarrow \text{top}(R_{remove})$;
$R_{add} \leftarrow \text{RankJMI}(U, S)$; $f_{add} \leftarrow \text{top}(R_{add})$;
$S_{minus} \leftarrow S \setminus \{f_{remove}\}$;
$S_{plus} \leftarrow S \cup \{f_{add}\}$;
$S_{swap} \leftarrow (S \setminus \{f_{remove}\}) \cup \{f_{add}\}$;
$m_{minus} \leftarrow \mathcal{M}(S_{minus})$; $m_{plus} \leftarrow \mathcal{M}(S_{plus})$; $m_{swap} \leftarrow \mathcal{M}(S_{swap})$;
$(S_{new}, m_{new}) \leftarrow$
$\arg\min\{(S, m), (S_{minus}, m_{minus}), (S_{plus}, m_{plus}), (S_{swap}, m_{swap})\}$;
if $m_{new} < m$ **then**
  $S \leftarrow S_{new}$; $m \leftarrow m_{new}$; $U \leftarrow \{1, \ldots, d\} \setminus S$;
  $\textit{patience} \leftarrow 0$;
else
  $\textit{patience} \leftarrow \textit{patience} + 1$;
if $m < m_{\text{best}}$ **then**
  $S_{\text{best}} \leftarrow S$; $m_{\text{best}} \leftarrow m$;
return $(S, U, m, S_{\text{best}}, m_{\text{best}}, \textit{patience})$;

Supplementary Experimental Information

Synthetic Dataset Design

The synthetic dataset utilized in this study is constructed with three balanced classes. The primary objective of this dataset is not to measure predictive performance, but rather to determine if relevant predictors are correctly identified by our CBFS within a controlled scenario.

Cohort Integration and Endpoint Harmonization

We integrated data from IMvigor210, IMvigor010, and ABACUS cohorts, which reflect different clinical settings (metastatic, adjuvant, and neoadjuvant) and utilize different time-to-event endpoints (PFS, DFS, and RFS). The rationale for integrating these cohorts is to investigate whether a baseline, treatment-related biological signal of atezolizumab benefit, detected by feature selection (FS), is shared across different disease stages. This analysis does not attempt to replicate the specific analyses of the original three studies, but rather compares the proposed method against established methods using real-world data.

To address endpoint heterogeneity, we harmonized the outcome based

on a shared clinical construct: event-free status at a fixed time horizon. Although PFS, DFS, and RFS differ in event definitions and follow-up contexts, all capture the concept of *lack of progression or recurrence* within a defined period. Consequently, the Durable Clinical Benefit at 6 months (DCB6) is defined identically across cohorts as *no event by 6 months*. This endpoint serves as a common binary surrogate for early durable benefit. To avoid ambiguous labeling, patients censored before the 6-month mark were excluded from the analysis, resulting in the removal of 115 patients due to censoring.

ABACUS Cohort Specifics for the leave-one-cohort-out validation

For the ABACUS cohort, the Cox validation model utilized recurrence-free survival (RFS) as the time-to-event endpoint. The follow-up time was defined by `RFS_months` and the event indicator by `RFS_event` (parsed as event=1 for Yes and censored=0 for No; entries coded as -1 and times < 0 were treated as missing). In this setting, an *event* corresponds to disease recurrence or death (loss of recurrence-free status), while censoring indicates no recorded event by the last follow-up.

This definition aligns directly with the DCB6 construction used in this study:

- **DCB6 = 1:** Patient is event-free at or beyond 6 months (RFS ≥ 6).
- **DCB6 = 0:** An event occurs before 6 months (event=1 and RFS < 6).

As noted previously, patients censored before 6 months are excluded to ensure label clarity.

Preprocessing and Validation Strategy

For data preprocessing, ComBat—a well-established method for genetic data—was applied to address batch effects, treating the cohort identifier as the batch variable. Variance filtering was employed to select the top 3,500 variable genes, utilizing variance as the variability metric.

Regarding the data splitting strategy, splits were performed after pooling the data, and cohort membership was strictly excluded from the feature selection stage. To assess generalizability and robustness, specifically regarding trial heterogeneity, a leave-one-cohort-out evaluation was conducted by

leaving the ABACUS cohort out of the training set. The detailed cohort summary table is provided in Table 1

Table 1: **Summary of Integrated Cohorts.** Overview of clinical settings, endpoints, and patient distribution across the three cohorts used for analysis.

Characteristic	**IMvigor210**	**IMvigor010**	**ABACUS**	**Total / Pooled**
Clinical Setting	Metastatic	Adjuvant	Neoadjuvant	—
Endpoint Definition	PFS	DFS	RFS	Event-free (6mo)
Original Sample Size (N)	208	728	148	1084
Excluded ($<$ 6mo Censored)	1	39	75	115
Final Analysis Size (N)	207	689	73	969
DCB6 Labels				
Responder (No Event $\geq$ 6mo)	149	479	69	606
Non-Responder (Event $<$ 6mo)	58	210	4	363

Note: PFS = Progression-Free Survival; DFS = Disease-Free Survival; RFS = Recurrence-Free Survival. DCB6 defined as event-free status at 6 months.