Uncertainty-Aware mRMR for Transferable Feature Selection in High-Dimensional, Low-Sample-Size Biomedical Data

Xuehan Chen XUC321@LEHIGH.EDU
Department of Computer Science and Engineering
Lehigh University
Bethlehem, PA, USA

Brian D. Davison BDD3@LEHIGH.EDU
Department of Computer Science and Engineering
Lehigh University
Bethlehem, PA, USA

Abstract

In biomedical and clinical research, high-dimensional data are often collected from multiple heterogeneous source groups, such as different hospitals, patient populations, cancer types, or experimental conditions. Researchers often seek a compact set of features, such as genes, that is informative across source groups and remains useful in newly collected, unseen target groups. Despite its practical importance, this *transferable feature selection* problem in the high-dimensional, low-sample-size (HDLSS) setting remains underexplored. In this work, we propose Uncertainty-Aware mRMR (UA-mRMR), a training-free filter method that extends the minimum Redundancy Maximum Relevance (mRMR) feature selection method by accounting for sample-size-dependent statistical uncertainty across heterogeneous source groups. Specifically, UA-mRMR uses Fisher z-transformed confidence bounds to stabilize correlation-based relevance and redundancy estimates under unequal sample sizes. Experiments on real-world multi-source biomedical datasets show that features selected by UA-mRMR transfer effectively to unseen target groups and achieve strong downstream predictive performance.

1. Introduction

Biomedical studies are increasingly conducted across multiple heterogeneous groups, such as different cancer types, or disease subgroups (Weinstein et al., 2013). In these settings, researchers often seek a compact set of features that captures target-relevant variation in the available source groups while remaining useful in newly collected target groups. This goal is particularly challenging in high-dimensional, low-sample-size (HDLSS) settings, where each group may contain only a limited number of samples and estimation reliability can vary substantially across groups. These challenges raise a key machine-learning question for healthcare: *How can we perform feature selection across heterogeneous source groups in HDLSS settings? Can accounting for sample-size uncertainty improve the transferability of selected features to unseen target groups?*

We study these challenges through feature selection techniques. Feature selection is particularly important in high-dimensional settings because it can reduce overfitting, improve interpretability, and reduce computational cost. Existing feature selection methods are commonly grouped into three categories: *filter*, *wrapper*, and *embedded* methods (Guyon and Elisseeff, 2003; Saeys et al., 2007). Filter methods rank features using model-agnostic criteria such as univariate statistical tests, mutual-information-based scores, Relief-style methods, or multivariate criteria such as mRMR (Saeys et al., 2007; Peng et al., 2005; Ding and Peng, 2005). Wrapper methods evaluate candidate feature subsets by repeatedly fitting a downstream predictor, often yielding strong task-specific performance but at substantial computational cost and with elevated risk of overfitting in small-sample regimes (Saeys et al., 2007). Embedded methods incorporate feature selection directly into model training such as the Lasso. In biomedical HDLSS problems, filter methods are especially attractive because they scale well to very high-dimensional data, remain classifier-agnostic, and preserve the interpretability of selected biomarkers (Saeys et al., 2007; Guyon and Elisseeff, 2003).

A related line of work addresses heterogeneity and distribution shift at the *predictor* level through domain generalization and distributionally robust optimization (Wang et al., 2022; Arjovsky et al., 2020; Duchi and Namkoong, 2021; Sagawa et al., 2020). These approaches aim to learn predictive models that remain robust across heterogeneous source environments and under distribution shift. While valuable for robust prediction, they do not directly address our *transferable feature selection* setting, where the goal is to identify a compact, interpretable, and transferable feature subset in a high-dimensional, low-sample-size multi-source biomedical setting. This distinction is important in healthcare applications, where selected features may themselves be scientifically or clinically meaningful, for example as candidate biomarkers or transferable signatures for downstream modeling on newly collected novel group datasets.

To study this problem, we focus on filter-based feature selection methods, which are especially attractive in high-dimensional, low-sample-size biomedical settings because they are computationally efficient, model-agnostic, and preserve the interpretability of selected variables. Among them, minimum Redundancy Maximum Relevance (mRMR) (Peng et al., 2005; Ding and Peng, 2005) provides a natural starting point: unlike simple univariate ranking methods, it selects features that are both individually informative about the target and minimally redundant with one another. Prior benchmark studies in high-dimensional biomedical data, including multi-omics settings, have also found that mRMR remains a competitive feature selection method and often performs well even when only a small number of features are selected (Li et al., 2022). However, standard mRMR does not explicitly account for the statistical uncertainty introduced by heterogeneous source groups with unequal sample sizes in the HDLSS setting.

When Pearson correlation is used to measure both relevance and redundancy in mRMR, multi-source datasets with small and imbalanced group sizes present an important challenge. In particular, correlation estimates from small groups can be unreliable because the sampling distribution of the Pearson correlation becomes highly variable when sample sizes are limited. Only when N is sufficiently large does this distribution become approximately normal, except in cases of extreme correlations. As a result, naive correlation-based feature

ranking may overemphasize unstable estimates from small groups and thereby compromise the generalizability of the selected feature sets to unseen groups or future datasets.

To address this challenge, we propose **Uncertainty-Aware mRMR (UA-mRMR)**, an extension of mRMR that accounts for sample-size uncertainty in correlation-based feature selection across heterogeneous source groups. We use the Fisher z-transformation (Fisher, 1921) to construct confidence intervals for correlation estimates of relevance and redundancy scores: lower confidence bounds (LCB) for per-group feature-target relevance and upper confidence bounds (UCB) for per-group feature-feature redundancy. This design reduces the influence of unstable estimates from small groups and yields feature subsets that transfer more reliably to unseen target groups.

Our contributions are as follows:

1. We formulate the underexplored problem of *transferable feature selection* in heterogeneous multi-source biomedical data, where selected features could be transferred from source groups to unseen target group. We identify a key challenge in this setting: standard correlation-based mRMR does not account for sample-size-dependent uncertainty under unequal group sizes, which can limit the transferability of selected feature subsets.

2. We propose **UA-mRMR**, an uncertainty-aware extension of mRMR-Corr that uses Fisher-z-based confidence bounds to obtain more reliable per-group relevance and redundancy estimates in the mRMR objective.

3. We evaluate UA-mRMR on TCGA across 15 cancer datasets and three omics modalities (CNV, RNA, miRNA) under a leave-one-group-out transfer protocol, comparing against eight filter, embedded, and distributionally robust baselines, and analyze performance across feature budgets and single-source vs. multi-source transfer setting.

Generalizable Insights about Machine Learning in the Context of Healthcare. Healthcare datasets are often assembled from heterogeneous cohorts of highly unequal size, such as hospitals, disease subtypes, while selected features must remain useful in newly collected cohorts. Our study suggests two broader lessons for healthcare machine learning in this setting. First, transferable feature selection in the HDLSS regime is affected by both cross-group heterogeneity and the reliability of group-specific estimates, with small groups contributing disproportionately noisy statistics. Second, explicitly accounting for sample-size-dependent uncertainty when aggregating group-level relevance and redundancy can improve the reliability and transferability of selected features while preserving interpretability. More broadly, this principle may be useful in other healthcare applications involving heterogeneous small cohorts, including biomarker discovery and small-sample medical imaging.

2. Related Work

Filter-based feature selection. Filter methods rank features using statistical criteria independent of any specific classifier (Guyon and Elisseeff, 2003; Chandrashekar and Sahin, 2014). Univariate filters such as the t-test and ReliefF (Urbanowicz et al., 2018) evaluate each feature independently, ignoring inter-feature dependencies. Information-theoretic

methods based on mutual information (MI) can capture nonlinear dependencies but are computationally expensive for large feature spaces (Vergara and Estévez, 2014). The mRMR framework (Peng et al., 2005; Ding and Peng, 2005) addresses this by greedily selecting features that maximize relevance to the target while minimizing redundancy with already-selected features. Brown et al. (2012) provide a unifying view showing that many MI-based criteria are special cases of conditional likelihood maximization. Our work extends the correlation-based mRMR variant by incorporating uncertainty quantification via Fisher z-transformations.

Embedded and wrapper methods. Lasso (Tibshirani, 1996; Fonti and Belitser, 2017) performs feature selection implicitly through ℓ_1-regularized regression. While effective, its selected features depend on the downstream model, and the regularization path can be sensitive to collinearity and sample size. Random forests (Breiman, 2001) offer built-in feature importance measures but are less suitable as standalone feature selectors in the HDLSS regime. Our approach is filter-based and thus model-agnostic, making it compatible with any downstream predictor.

Distributionally robust feature selection. Recent work has applied distributionally robust optimization (DRO) to feature selection to improve worst-case performance across subpopulations (Sagawa et al., 2020; Duchi and Namkoong, 2021). DRO-Lasso (Chen and Paschalidis, 2022) extends grouped variable selection with robustness guarantees, while DROFS (Swaroop et al., 2025) optimizes feature subsets for distributional robustness. However, these methods are typically evaluated on in-domain test sets. We find that in the HDLSS setting, where features are selected on source data and evaluated on an unseen target domain, DRO-based methods can underperform simpler filter baselines.

3. Methods

3.1. Problem Setup

We consider transferable feature selection in the multi-source, high-dimensional, low-sample-size (HDLSS) regime. Let

$$\mathcal{D}_{\text{src}} = \{D_1, D_2, \ldots, D_G\}$$

denote G source groups, where each group

$$D_g = \{(\mathbf{x}_i^{(g)}, y_i^{(g)})\}_{i=1}^{n_g}, \qquad \mathbf{x}_i^{(g)} \in \mathbb{R}^p, \quad y_i^{(g)} \in \{0, 1\},$$

contains n_g samples with p-dimensional feature vectors and binary labels. We focus on the regime $p \gg n_g$, with potentially large imbalance among group sizes n_g. Given a feature budget $K \ll p$, our goal is to select a subset

$$S \subseteq \{1, \ldots, p\}, \qquad |S| = K,$$

using only the source groups $\mathcal{D}_{\text{src}}$, such that the selected features remain predictive when transferred to a previously unseen target group D_{tgt} drawn from a related but distinct distribution. In contrast to within-dataset feature selection, the criterion of success is not in-source predictive utility but *transferability*: a feature subset is useful only if it generalizes beyond the source groups used for selection.

In biomedical applications, the datasets correspond to natural data partitions such as hospitals or clinical sites (Bandi et al., 2018), cancer types (Weinstein et al., 2013), sequencing batches, or patient demographic strata. These groups are typically high-dimensional, individually small, distributionally heterogeneous, and unequal in size: in our TCGA experiments, BRCA contains ∼970 samples while ESCA contains only ∼140 (Li et al., 2022). A clinically useful biomarker panel need remain informative when a new group is profiled. UA-mRMR is designed for this multi-source setting, which is our primary focus.

3.2. Maximize Relevance and Minimize redundancy (mRMR)

The mRMR framework (Peng et al., 2005; Ding and Peng, 2005) selects features sequentially by maximizing relevance to the label while minimizing redundancy with already selected features. At each step, it adds the feature

$$j^* \;=\; \arg\max_{j \notin S} \left[\mathrm{Rel}(j) - \frac{1}{|S|} \sum_{j' \in S} \mathrm{Red}(j, j') \right] \tag{1}$$

to the current selected set S, where $\mathrm{Rel}(j)$ denotes the *relevance* of feature j by quantifying its dependence on the label y ,and $\mathrm{Red}(j, j')$ denotes the *redundancy* between features j and j' by quantifying their relationship. Different choices of these two quantities give rise to different mRMR variants.

Mutual Information Difference (mRMR-MID). The original mRMR formulation uses mutual information for both relevance and redundancy (Peng et al., 2005):

$$\mathrm{Rel}(j) \;=\; I(X_j;\, y), \qquad \mathrm{Red}(j, j') \;=\; I(X_j;\, X_{j'}). \tag{2}$$

Here,

$$I(X;\, Y) \;=\; \iint p(x, y)\, \log \frac{p(x, y)}{p(x)\, p(y)}\, dx\, dy. \tag{3}$$

Mutual information captures nonlinear dependence, but estimating it reliably in the HDLSS regime is difficult because it requires density estimation (Vergara and Estévez, 2014).

F-test Correlation Difference (mRMR-FCD). For continuous features, Ding and Peng (2005) proposed replacing mutual-information relevance with the one-way ANOVA F-statistic and redundancy with absolute Pearson correlation:

$$\mathrm{Rel}(j) \;=\; F(X_j;\, y), \qquad \mathrm{Red}(j, j') \;=\; |\hat{\rho}(X_j,\, X_{j'})|. \tag{4}$$

For K classes $\{\mathcal{C}_k\}$ of sizes $\{n_k\}$,

$$F(X_j;\, y) \;=\; \frac{\displaystyle\sum_{k=1}^{K} n_k (\bar{x}_{jk} - \bar{x}_j)^2 / (K-1)}{\displaystyle\sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} (x_{ij} - \bar{x}_{jk})^2 / (n-K)}, \tag{5}$$

and

$$\hat{\rho}(X_j,\, X_{j'}) \;=\; \frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ij'}-\bar{x}_{j'})}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2\sum_{i=1}^{n}(x_{ij'}-\bar{x}_{j'})^2}}. \tag{6}$$

Correlation-based mRMR (mRMR-Corr). We also consider using absolute Pearson correlation for both relevance and redundancy as one variant of mRMR method:

$$\mathrm{Rel}(j) \;=\; |\hat{\rho}(X_j,\, y)|, \qquad \mathrm{Red}(j,j') \;=\; |\hat{\rho}(X_j,\, X_{j'})|. \tag{7}$$

This choice is motivated by prior correlation-based extensions of mRMR, which use Pearson correlation as a redundancy measure and correlation-style relevance measures in place of mutual information (Jo et al., 2019). Since y is binary in our setting, $|\hat{\rho}(X_j, y)|$ is equivalent to the absolute point-biserial correlation between feature X_j and the label.

Limitations in the multi-source HDLSS regime. In the multi-source HDLSS setting, feature dependence is estimated separately within groups whose sample sizes may differ substantially. A key difficulty is that sample correlations from small groups are much noisier than those from larger groups, and their uncertainty also depends on the unknown true correlation. As a result, raw correlation-based relevance or redundancy estimates are not directly comparable across groups, which can make feature ranking sensitive to small-sample fluctuations rather than cross group signals.

3.3. Uncertainty-Aware mRMR

To address transferable feature selection under imbalance and small sample size, we propose Uncertainty-Aware mRMR (UA-mRMR), a group-aware variant of correlation-based mRMR that incorporates sample-size uncertainty through Fisher z confidence bounds. The key idea is to compute relevance and redundancy within each source group, adjust these estimates conservatively according to their group-specific uncertainty, and then aggregate them across groups by uniform averaging.

Fisher z-based uncertainty estimation. For a sample correlation $\hat{\rho}_g$ computed from n_g observations in group g, the Fisher z-transformation (Fisher, 1921) is

$$z_g = \operatorname{arctanh}(\hat{\rho}_g) = \tfrac{1}{2}\ln\!\left(\frac{1+\hat{\rho}_g}{1-\hat{\rho}_g}\right). \tag{8}$$

Under the standard Fisher approximation, z_g is approximately Gaussian:

$$z_g \dot{\sim} \mathcal{N}(\mu_g, \sigma_g^2), \tag{9}$$

with

$$\mu_g \approx \operatorname{arctanh}(\rho_g), \tag{10}$$

$$\sigma_g^2 \approx \frac{1}{n_g-3}, \tag{11}$$

where ρ_g denotes the population correlation in group g.

For each group g, feature j, and feature pair (j, j'), we define the within-group empirical correlations and their Fisher z transforms as

$$\hat{\rho}_g(j) := |\widehat{\mathrm{corr}}(X_j^{(g)}, y^{(g)})|, \qquad z_g(j, y) := \operatorname{arctanh}\big(\hat{\rho}_g(j)\big),$$

$$\hat{\rho}_g(j, j') := |\widehat{\mathrm{corr}}(X_j^{(g)}, X_{j'}^{(g)})|, \qquad z_g(j, j') := \operatorname{arctanh}\big(\hat{\rho}_g(j, j')\big).$$

We restrict attention to groups with sufficient samples for Fisher z to be well-defined,

$$\mathcal{G} := \{g : n_g \geq 4\},$$

and write $\sigma_g = 1/\sqrt{n_g - 3}$ for the group-specific Fisher standard error.

Uncertainty-aware relevance and redundancy. To avoid overstating relevance in small groups with spuriously large feature–label correlations, we apply a lower-bound adjustment to each group-specific relevance estimate and then transform back to the correlation scale before averaging:

$$\widetilde{\mathrm{Rel}}(j) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \tanh\Big(\max\big(0,\ z_g(j, y) - c\,\sigma_g\big)\Big). \tag{12}$$

This downweights group contributions according to their uncertainty, with smaller groups receiving more conservative relevance estimates.

To avoid understating redundancy in small groups with spuriously weak feature–feature correlations, we apply an upper-bound adjustment:

$$\widetilde{\mathrm{Red}}(j, j') = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \tanh\big(z_g(j, j') + c\,\sigma_g\big). \tag{13}$$

This increases redundancy estimates for groups whose dependence structure is more uncertain, again penalizing smaller groups more strongly.

We use a lower bound for relevance and an upper bound for redundancy because, in the mRMR objective, relevance acts as a benefit term whereas redundancy acts as a cost term. Under estimation uncertainty, a conservative selection rule should avoid overestimating benefits and underestimating costs. Accordingly, we discount potentially overstated relevance and increase potentially understated redundancy, thereby favoring stable and transferable signals over correlations driven by small-sample noise.

The scalar $c \geq 0$ sets the bound width in Fisher-z standard errors ($c = 1$ is the nominal $\approx 68\%$ interval, $c = 1.96$ the 95% interval); we use a one standard error ($c = 1$) in our experiment. A sensitivity analysis in Appendix B.3, Table 5 shows the ranking is stable across c and peaks near 1.

The final selection rule has the same form as standard mRMR, but replaces the original relevance and redundancy terms with the uncertainty-aware quantities in Eqs. (12) and (13):

$$j^* \;=\; \arg\max_{j \notin S} \Big[\widetilde{\mathrm{Rel}}(j) \;-\; \frac{1}{|S|} \sum_{j' \in S} \widetilde{\mathrm{Red}}(j, j') \Big]. \tag{14}$$

Algorithm 1: UA-mRMR feature selection

Input: Source groups $\{(\mathbf{X}^{(g)}, \mathbf{y}^{(g)})\}_{g=1}^{G}$, feature budget K
Output: Selected feature subset S with $|S| = K$

$\mathcal{G} \leftarrow \{g : n_g \geq 4\}$; $S \leftarrow \emptyset$;
foreach $g \in \mathcal{G}$ **do**
  $\hat{\rho}_g(j) \leftarrow |\widehat{\mathrm{corr}}(X_j^{(g)}, y^{(g)})|$ for all $j \in \{1, \ldots, p\}$;
  $\hat{\rho}_g(j, j') \leftarrow |\widehat{\mathrm{corr}}(X_j^{(g)}, X_{j'}^{(g)})|$ for all (j, j');
  $z_g(j) \leftarrow \mathrm{arctanh}(\hat{\rho}_g(j))$; $z_g(j, j') \leftarrow \mathrm{arctanh}(\hat{\rho}_g(j, j'))$;
end

$$\widetilde{\mathrm{Rel}}(j) \leftarrow \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \tanh\Big(\max(0,\ z_g(j) - 1/\sqrt{n_g - 3})\Big) \quad \text{for all } j;$$

$$\widetilde{\mathrm{Red}}(j, j') \leftarrow \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \tanh\Big(z_g(j, j') + 1/\sqrt{n_g - 3}\Big) \quad \text{for all } (j, j');$$

$S \leftarrow \big\{\arg\max_j \widetilde{\mathrm{Rel}}(j)\big\}$;
for $t \leftarrow 2$ **to** K **do**
  $$j^* \leftarrow \arg\max_{j \notin S} \Big[\widetilde{\mathrm{Rel}}(j) - \frac{1}{|S|} \sum_{j' \in S} \widetilde{\mathrm{Red}}(j, j')\Big];$$
  $S \leftarrow S \cup \{j^*\}$;
end
return S;

Selection is initialized by

$$S \leftarrow \left\{ \arg\max_j \widetilde{\mathrm{Rel}}(j) \right\},$$

and Eq. (14) is iterated until $|S| = K$.

Computational cost. Computationally, the dominant cost lies in computing the per-group feature–feature correlation matrices, which requires $O(p^2 n_g)$ per group and $O(p^2 N)$ overall. Once relevance and redundancy are precomputed, the greedy selection step costs $O(Kp)$. The method is training-free, as it relies only on group-wise summary statistics and does not fit a predictive model during selection. Compared with embedded methods that require iterative optimization, UA-mRMR avoids model training during feature selection, typically reducing computational overhead and directly producing a feature subset of the desired size.

3.4. Transferable Feature Selection Framework

A central goal of this work is to evaluate whether selected features remain useful when transferred to a newly profiled group that was not available during selection. We hypothesize that uncertainty-aware aggregation across heterogeneous source groups produces feature subsets whose predictive utility depends less on group-specific noise and more on the cross-group signal, and is therefore more transferable to unseen groups. To test this hypothesis,

we design a multi-source transferable feature selection framework based on leave-one-group-out evaluation, which is suitable for the heterogeneous, multi-group HDLSS setting that motivates UA-mRMR.

Let $\mathcal{D} = \{D_1, \ldots, D_M\}$ denote the full collection of groups available for evaluation. In each round, one group D_t is held out as the *target* group, and the remaining groups

$$\mathcal{D}_{\text{src}}^{(t)} := \mathcal{D} \setminus \{D_t\}$$

serve as *source* groups. Feature selection is performed using only $\mathcal{D}_{\text{src}}^{(t)}$, without any access to D_t. The resulting subset is then evaluated on the held-out target group by retraining a downstream classifier on the target's training portion using only the selected features and reporting test performance on the target's held-out portion. This protocol evaluates the transferability of the selected features rather than that of the downstream classifier: the classifier may adapt to the target data, but only within the feature space selected from the source groups, so differences in target performance can be attributed primarily to the quality of the feature selection step.

The framework supports two complementary evaluation modes that distinguish how source information is aggregated.

Multi-source transfer (primary setting). A shared feature subset S^{multi} is selected jointly from all source groups in $\mathcal{D}_{\text{src}}^{(t)}$ and then transferred to the target group D_t. On D_t, the downstream classifier is retrained and evaluated using only the features in S^{multi}. This setting tests whether jointly leveraging multiple heterogeneous source groups leads to a more transferable feature subset. UA-mRMR is specifically designed for this setting.

Single-source transfer (complementary setting). For each source group $D_s \in \mathcal{D}_{\text{src}}^{(t)}$, an independent feature subset S_s is selected using only D_s and then transferred to the target group D_t, where the downstream classifier is retrained using only the features in S_s. Performance on the target group is then averaged across source groups. This setting measures how well features selected from a single source group transfer to an unseen target group, and serves as a baseline for quantifying the benefit of joint aggregation in the multi-source setting.

This procedure is repeated over all possible target groups, and the final performance is averaged across targets. We report downstream test AUROC as the primary metric. Unlike conventional feature selection benchmarks, this framework never evaluates performance on the same group used for selection, but rather on transfer to a previously unseen group, which is the practically relevant question in clinical biomarker discovery.

4. Experiments

4.1. Datasets

We use data from The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013), focusing on 15 diseases: BLCA, BRCA, COAD, ESCA, HNSC, LGG, LIHC, LUAD, LUSC, PAAD, PRAD, SARC, SKCM, STAD, and UCEC. This collection is well suited to our experimental setting, which evaluates whether feature sets selected by different feature selection methods

generalize to unseen target datasets. The main characteristics of these datasets are summarized in Table 2 in Appendix A. We use *TP53* mutation status as the prediction target, following prior work that treats it as a clinically relevant surrogate phenotype due to its association with poor cancer outcomes (Li et al., 2022).

To assess how well the proposed methods generalize across different feature distributions, we extracted data from three distinct genomic modalities for these 15 diseases:

- **TCGA-CNV**: Copy number variation profiles with 57,408 features per sample.
- **TCGA-RNA**: RNA expression profiles with 17,355 features per sample.
- **TCGA-miRNA**: miRNA expression profiles with 495 features per sample.

Data preprocessing. To ensure a common feature space within each modality, we first retained only the features present across all 15 cancer types. For high-dimensional modalities ($p > 500$), we applied a variance-based pre-filter separately within each leave-one-disease-out fold: after holding out target disease D_t, we computed feature variances using only the pooled source data $D_{\text{src}}^{(t)} = D \setminus \{D_t\}$, retained the top 500 high-variance features, and then applied this source-defined feature subset to the held-out target. Thus, the target disease did not influence feature filtering or any other preprocessing step prior to evaluation. The variance pre-filter applies only where $p > 500$, i.e. to TCGA-CNV (57,408 $\rightarrow$ 500) and TCGA-RNA (17,355 $\rightarrow$ 500); TCGA-miRNA has 495 common features and is therefore evaluated on the full feature space for all methods, an asymmetry worth noting when comparing modality-specific patterns. We fix the feature size at 500 as a tractable midpoint for the $O(p^2)$ mRMR redundancy; the UA-mRMR versus mRMR-Corr ranking is stable across caps $\{100, 500, 1000\}$ (see Appendix B.4), so the comparison reflects the selection method rather than the variance filter.

4.2. Experimental Setup

We evaluate all methods under the transferable feature selection framework described in Section 3.4 using leave-one-group-out transfer. For each modality, performance is reported as mean $\pm$ standard deviation AUROC across the 15 cancer datasets, averaged over 5 random seeds and two downstream classifiers. The downstream classifiers are **Logistic Regression (LR)**, with regularization parameter $C \in \{0.1, 1.0, 10.0\}$ selected by 3-fold stratified cross-validation, and **Random Forest (RF)** (Breiman, 2001), with 100 trees and maximum depth in $5, 10$ selected by 3-fold stratified cross-validation. Feature budgets are $K \in \{5, 10, 15, 20, 25, 30\}$ for all methods.

4.3. Baseline Methods

We compare UA-mRMR against eight standard and competitive baselines spanning filter, embedded, and distributionally robust feature selection. All methods are evaluated under the same transferable feature selection setting in order to assess how well the selected feature subsets generalize to an unseen target group.

- **t-test**: This univariate baseline ranks features by the two-sample t-statistic between classes. It serves as a simple relevance-based method without any mechanism to control feature redundancy.

- **ReliefF** (Urbanowicz et al., 2018): ReliefF estimates feature importance by comparing each sample with its nearest neighbors from the same and different classes. It is included as a widely used relevance-based method that can capture local discriminative structure.
- **mRMR-MI** (Peng et al., 2005): mRMR method uses mutual information to measure both feature–target relevance and feature–feature redundancy.
- **mRMR-FCD** (Ding and Peng, 2005): This mRMR variant uses the F-statistic to measure relevance and absolute Pearson correlation to measure redundancy.
- **mRMR-Corr**: This variant uses absolute Pearson correlation for both relevance and redundancy. It is the most direct baseline for our method, since UA-mRMR is built by introducing uncertainty-aware adjustments on top of correlation-based mRMR.
- **Lasso** (Tibshirani, 1996): We use Lasso as an embedded feature selection baseline implemented through ℓ_1-regularized logistic regression. It selects features by shrinking less informative coefficients to zero during model fitting.
- **DRO-Lasso** (Chen and Paschalidis, 2022): DRO-Lasso extends Lasso with a distributionally robust optimization objective that minimizes the worst-case loss across observed groups.
- **DROFS** (Swaroop et al., 2025): DROFS is a distributionally robust feature selection method that explicitly accounts for group structure during selection. Instead of selecting features solely to maximize average predictive performance, it optimizes for robustness across observed subpopulations, such as institutions or demographic groups.

Although DRO-Lasso and DROFS are designed to improve robustness across observed groups rather than explicitly optimize transferability to a completely unseen target group, we include them as baselines to evaluate how well such robust feature selection methods perform under our proposed transferable feature selection framework.

5. Results

We first evaluate the multi-source transfer performance of all methods across the three TCGA modalities (CNV, RNA, and miRNA) at $K = 25$, averaging results over 15 unseen target cancer types, 5 random seeds, and both downstream classifiers (Table 1).

Among the filter methods, mRMR-Corr achieves the best AUROC (LR: 0.673, RF: 0.707), whereas t-test and ReliefF perform less strongly, suggesting that redundancy-aware multivariate selection improves transferability. Within the mRMR family, mRMR-Corr performs best, followed by mRMR-FCD and then mRMR-MI, indicating that correlation-based estimators are better suited than mutual information to HDLSS data with continuous, correlated features. Among the embedded methods, Lasso remains competitive (LR AUROC: 0.681, RF AUROC: 0.713) despite requiring model fitting, and DRO-Lasso is similarly strong, whereas DROFS underperforms, suggesting its distributionally robust objective is less well aligned with the unseen-target transfer scenario. UA-mRMR achieves the

highest AUROC under both classifiers (0.694 with LR and 0.714 with RF), outperforming all baselines including the embedded Lasso while remaining training-free, indicating that uncertainty-aware per-group aggregation yields more transferable feature subsets.

Category	Method	Logistic Regression		Random Forest	
		Accuracy	**AUROC**	**Accuracy**	**AUROC**
Filter	t-test	0.720 ± 0.095	0.667 ± 0.091	0.731 ± 0.095	0.690 ± 0.097
	ReliefF	0.714 ± 0.092	0.663 ± 0.076	0.736 ± 0.083	0.695 ± 0.095
mRMR variants	mRMR-MI	0.715 ± 0.090	0.648 ± 0.093	0.732 ± 0.088	0.677 ± 0.100
	mRMR-FCD	0.720 ± 0.095	0.665 ± 0.092	0.729 ± 0.096	0.686 ± 0.097
	mRMR-Corr	0.724 ± 0.085	0.673 ± 0.100	0.746 ± 0.083	0.707 ± 0.104
Embedded	Lasso	0.718 ± 0.091	0.681 ± 0.091	0.751 ± 0.080	0.713 ± 0.097
Distrib. Robust	DRO-Lasso	0.721 ± 0.088	0.677 ± 0.086	0.743 ± 0.080	0.702 ± 0.099
	DROFS	0.711 ± 0.092	0.644 ± 0.096	0.729 ± 0.091	0.678 ± 0.102
Proposed	UA-mRMR	**0.724 ± 0.089**	**0.694 ± 0.078**	**0.752 ± 0.078**	**0.714 ± 0.101**

Table 1: Cross-group transfer performance at $K = 25$. For each method and each target cancer type, AUROC and accuracy are first averaged over 5 random seeds and 2 downstream classifiers within each TCGA modality. These modality-specific values are then averaged over the 3 TCGA modalities (CNV, RNA, and miRNA).

While Table 1 evaluates all methods at a single fixed budget K=25, the choice of budget is application-dependent, making it important to understand how methods behave across a wider range. We therefore compare all mRMR-family methods over feature budgets $K \in \{5, 10, 15, 20, 25, 30\}$ to assess whether UA-mRMR's advantage is consistent across different datasets.

Figure 1 reveals three consistent patterns. First, UA-mRMR outperforms all other mRMR variants across most feature budgets K, so its advantage is not specific to the K=25 setting of Table 1. Second, the gap between UA-mRMR and mRMR-Corr tends to widen as K increases on TCGA-RNA and TCGA-miRNA, suggesting that adjusting per-group correlation estimates for sample-size uncertainty becomes increasingly beneficial when more features are selected. Third, mRMR-FCD outperforms mRMR-MI across all modalities and nearly all budgets (largest gap on TCGA-CNV), likely because mutual information is harder to estimate reliably with limited samples.

Figure 2 shows the per-target difference between UA-mRMR and mRMR-Corr when each cancer dataset is treated in turn as an unseen target. Across the three modalities, UA-mRMR improves over mRMR-Corr on 9/15 targets for CNV and 10/15 targets for both RNA and miRNA, showing that modeling sample-size uncertainty often yields more transferable feature subsets than mRMR-Corr.

So far we have focused on multi-source transfer, where feature selection uses all available source cancer datasets jointly. In practice, a researcher may have access to only a single source dataset and wish to transfer the selected features to another disease dataset; we therefore compare single-source and multi-source transfer for the mRMR-based methods. Because single-source evaluation over all source–target pairs is substantially more expensive,

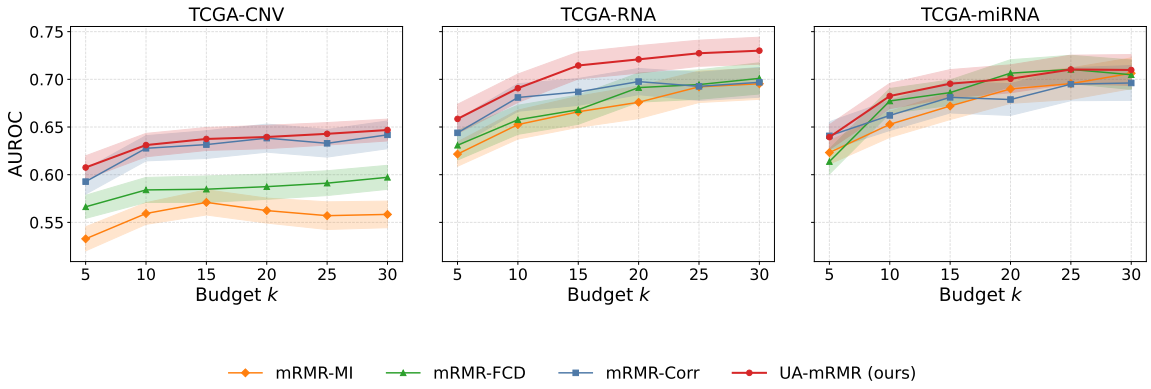

Figure 1: AUROC of mRMR-family methods (Logistic Regression) across feature budgets $K \in \{5, \ldots, 30\}$ on the three TCGA modalities under multi-source transfer. Lines show the mean over 15 unseen target cancer types $\times$ 5 seeds = 75 evaluations per point; shaded bands show ± 1 standard error of the mean. UA-mRMR (red) is above other mRMR variants across most budget and modality.

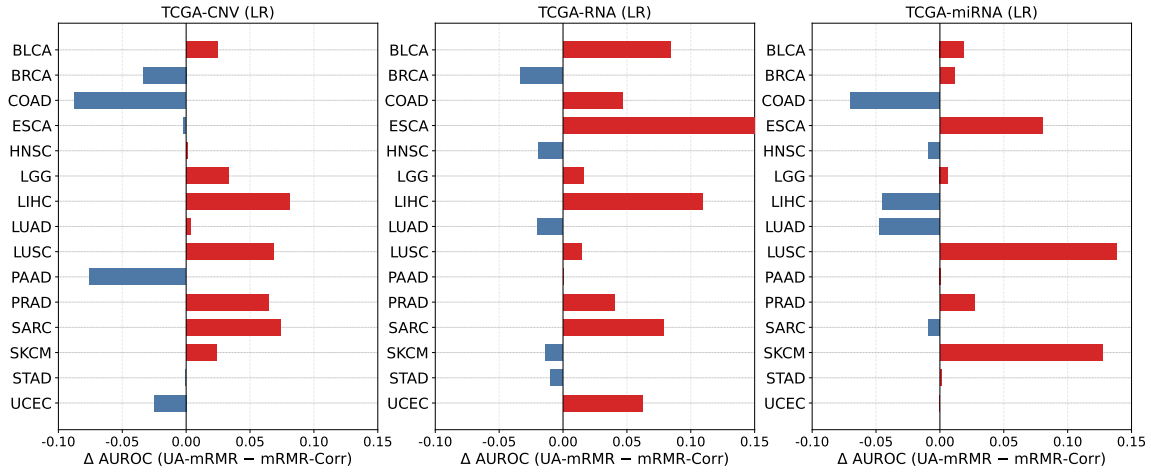

Figure 2: Per-target ΔAUROC at K=25 (UA-mRMR $-$ mRMR-Corr) for each of the 15 held-out cancer types across the three TCGA modalities (LR only, averaged over 5 seeds). For each target cancer type, the feature subset is selected using the other 14 cancer types as source datasets and then evaluated on the held-out target. Red bars indicate targets where UA-mRMR outperforms mRMR-Corr, while blue bars indicate the opposite.

we restrict this comparison to mRMR-FCD, mRMR-Corr (the two stronger filter baselines), and UA-mRMR.

For both mRMR-FCD and mRMR-Corr, multi-source selection yields higher mean AUROC than single-source selection (Figure 3), indicating that jointly using multiple source

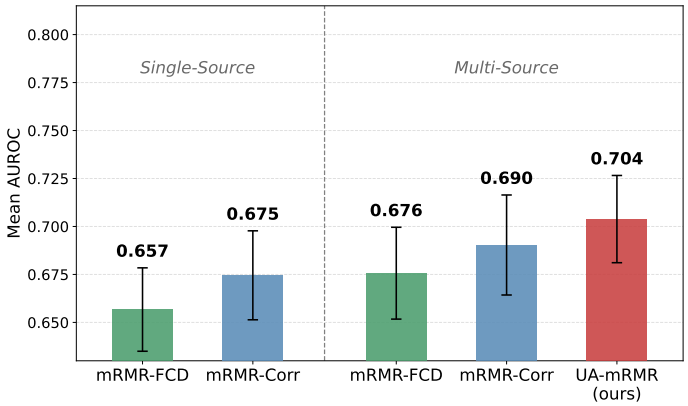

Figure 3: Single-source vs. multi-source transfer at K=25, averaged over LR and RF and the three TCGA modalities. In the single-source setting, features are selected independently on each source dataset and then transferred to target datasets. In the multi-source setting, features are selected jointly from all source datasets before transfer.

datasets leads to more transferable feature subsets, and UA-mRMR further improves over both baselines to achieve the best overall performance.

6. Discussion

Why does UA-mRMR help? Transferability in HDLSS multi-source data faces two distinct challenges: feature–label correlations estimated within small groups are unreliable (*within-group estimation noise*; e.g. ESCA $n \approx 140$ vs. BRCA $n \approx 970$), and true associations differ across cancer types (*between-group heterogeneity*). UA-mRMR addresses the second through group-wise estimation and aggregation and the first through Fisher-z bounds: applying LCB to relevance retains a feature only when its target association has reliable cross-group support, while applying UCB to redundancy penalizes features that appear non-redundant only because small-group redundancy estimates are noisy. UA-mRMR thereby favors feature subsets that transfer better to unseen targets.

Why do DRO methods underperform in transfer feature selection? DROFS achieves the lowest AUROC among the baselines despite being designed for distributional robustness. We also tuned its regularization strength $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ by internal validation, and it still underperformed UA-mRMR. This is likely because DRO methods optimize worst-case performance across *source* subpopulations, which is not equivalent to generalizing to an *unseen target*: worst-group optimization may favor features that perform evenly across observed sources but are not the most transferable. By contrast, UA-mRMR down-weights unreliable small-group correlation estimates, so the final ranking reflects more trustworthy cross-group signal.

Limitations. Several limitations should be acknowledged:

1. **Single prediction target on TCGA.** All TCGA experiments use *TP53* mutation status as the target. While *TP53* is a clinically relevant surrogate phenotype associated with poor cancer outcomes, it does not cover the full range of clinically important prediction tasks. Generalization to other targets, such as survival and treatment response, remains future work. We additionally evaluated a clinically distinct target, advanced pathologic stage (Stage III/IV), and present results in Appendix C; the advantage carries over on CNV but narrows on this harder, near-chance target, where all methods are close.

2. **Gaussian assumption.** Our uncertainty calibration relies on the Fisher-z approximation, which is derived under approximate Gaussian assumptions. Although many real-world biomedical features deviate from normality, especially in HDLSS settings, this approximation still provides a practical way to account for sample-size uncertainty.

7. Conclusion

We formulated *transferable feature selection* in heterogeneous multi-source HDLSS biomedical data and proposed UA-mRMR as a training-free filter method for this setting. By combining group-wise correlation estimation with Fisher-z confidence bounds, UA-mRMR reduces the influence of unreliable estimates from small source groups and selects feature subsets that transfer more reliably to unseen targets. Under a leave-one-group-out evaluation framework covering both single-source and multi-source transfer on 15 TCGA cancer types across three modalities, UA-mRMR achieves the strongest overall performance against filter, embedded, and distributionally robust baselines, with the highest average AUROC and accuracy, outperforming baselines at most feature budgets. Ablation and significance analyses further show that group-aware averaging, Fisher-z uncertainty adjustment, and their asymmetric use on relevance and redundancy each contribute to the final gain. Overall, jointly accounting for source-group heterogeneity and small-sample uncertainty is key to selecting transferable features in HDLSS biomedical data.

Acknowledgments

This research is supported by funding from the NIH grant 1R01EB033651-01A1 (Nanosensor Array Platform to Capture Whole Disease Fingerprints) and NSF grant 2323759 (DMREF: DNA-Nanocarbon Hybrid Materials for Perception-Based, Analyte-Agnostic Sensing).

References

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization, 2020. URL https://arxiv.org/abs/1907.02893.

Peter Bandi, Oscar Geessink, Quirine Manber, Marcory van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Transactions on Medical Imaging*, 38(2):550–560, 2018.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Luján. Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13:27–66, 2012.

Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

Ruidi Chen and Ioannis Ch Paschalidis. Robust grouped variable selection using distributionally robust optimization. *Journal of Optimization Theory and Applications*, 194(3): 1042–1071, 2022.

Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(02):185–205, 2005.

John C. Duchi and Hongseok Namkoong. Learning models with uniform performance via distributionally robust optimization. *The Annals of Statistics*, 49(3):1378–1406, 2021.

Ronald A Fisher. On the" probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32, 1921.

Valeria Fonti and Eduard Belitser. Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 30:1–25, 2017.

Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

Insik Jo, Sangbum Lee, and Sejong Oh. Improved measures of redundancy and relevance for mrmr feature selection. *Computers*, 8(2):42, 2019.

Yingxia Li, Ulrich Mansmann, Shangming Du, and Roman Hornung. Benchmark study of feature selection strategies for multi-omics data. *BMC bioinformatics*, 23(1):412, 2022.

Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731*, 2020.

Maitreyi Swaroop, Tamar Krishnamurti, and Bryan Wilder. Distributionally robust feature selection. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288, 1996.

Ryan J. Urbanowicz, Melissa Meeker, William LaCava, Randal S. Olson, and Jason H. Moore. Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, 85:189–203, 2018.

Jorge R. Vergara and Pablo A. Estévez. A review of feature selection methods based on mutual information. *Neural Computing and Applications*, 24(1):175–186, 2014.

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip S. Yu. Generalizing to unseen domains: A survey on domain generalization, 2022. URL https://arxiv.org/abs/2103.03097.

John N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna R. Mills Shaw, Brad A. Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M. Stuart. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113–1120, 2013.

Appendix A. Dataset Characteristics

Dataset	Cancer	n	Pos	Neg	r_{Pos}
BRCA	Breast invasive carcinoma	735	255	480	0.347
HNSC	Head and neck squamous cell carcinoma	443	307	136	0.693
LUAD	Lung adenocarcinoma	426	212	214	0.498
LGG	Brain lower grade glioma	419	195	224	0.465
LUSC	Lung squamous cell carcinoma	418	346	72	0.828
PRAD	Prostate adenocarcinoma	407	48	359	0.118
UCEC	Uterine corpus endometrial carcinoma	405	144	261	0.356
BLCA	Bladder urothelial carcinoma	382	186	196	0.487
STAD	Stomach adenocarcinoma	295	139	156	0.471
SKCM	Skin cutaneous melanoma	249	39	210	0.157
COAD	Colon adenocarcinoma	191	106	85	0.555
LIHC	Liver hepatocellular carcinoma	159	44	115	0.277
SARC	Sarcoma	126	48	78	0.381
PAAD	Pancreatic adenocarcinoma	124	78	46	0.629
ESCA	Esophageal carcinoma	106	83	23	0.783

Table 2: Data characteristics for 15 diseases. The second column lists the full cancer type name. The last four columns report the total number of samples (n), the number of *TP53*-mutant cases (Pos), the number of *TP53* wild-type cases (Neg), and the the ratio between the numbers of mutation events and the numbers of observations (r_{Pos}).

Appendix B. Ablation Studies

To evaluate the contribution of each component in our proposed framework, we report a set of ablations and sensitivity analyses under the same multi-source leave-one-group-out protocol as Table 1, fixing the feature budget at K=25 and using both Logistic Regression (LR) and Random Forest (RF) downstream classifiers, averaged over 5 random seeds and the 15 unseen target cancer types.

B.1. Decomposition: Group-Aware Averaging vs. Fisher-z Uncertainty

We conduct an ablation study to measure the performance gains that are directly due to our two primary contributions. *(i)* the group-aware averaging mechanism and *(ii)* the uncertainty-driven Fisher-z adjustment. This analysis allows us to verify whether accounting for small-sample uncertainty via the Fisher-z transformation provides benefits beyond simple uniform group-averaging. We compare three variants:

- **mRMR-Corr:** the correlation-based baseline used in the main paper, which computes a single correlation for relevance and redundancy on the union of all source samples without considering group structure for diseases.

- **UA-mRMR (w/o Fisher-z):** a group-aware version of mRMR-Corr in which relevance and redundancy are computed within each source group and averaged uniformly across groups. No Fisher-z confidence adjustment is applied; this variant therefore isolates the effect of group-aware averaging alone.
- **UA-mRMR:** the full design, which additionally applies the Fisher-z lower and upper confidence bounds on top of the per-group averages, trimming correlations estimated from smaller, noisier source groups.

Method	Logistic Regression		Random Forest	
	Accuracy	**AUROC**	**Accuracy**	**AUROC**
mRMR-Corr	0.737 ± 0.087	0.693 ± 0.110	0.744 ± 0.091	0.696 ± 0.121
UA-mRMR (w/o Fisher-z)	0.755 ± 0.085	0.725 ± 0.099	0.755 ± 0.092	0.702 ± 0.119
UA-mRMR	**0.756 ± 0.080**	**0.731 ± 0.093**	**0.758 ± 0.087**	**0.704 ± 0.117**

Table 3: Contribution of Group Aggregation and Fisher-z Adjustment in UA-mRMR. Values are mean ± standard deviation across 15 unseen target cancer types for TCGA RNA modality.

As shown in Table 3, moving from mRMR-Corr to UA-mRMR (w/o Fisher-z) yields an average absolute improvement of 0.0168 across downstream metrics, and adding the Fisher-z adjustment further improves performance by 0.0030 on average. While these mean-level gains are relatively small, they indicate that both group-aware aggregation and uncertainty-aware adjustment are beneficial.

B.2. Relevance and Redundancy Adjustment

To evaluate the individual contributions of the relevance and redundancy adjustments, we perform an ablation study comparing the full UA-mRMR objective against two partial variants and the mRMR-Corr baseline:

- **UA-mRMR (LCB):** a variant in which the Fisher-z-based lower confidence bound is applied only to the per-group relevance term, while redundancy is computed using the standard $|\mathrm{corr}(X_j, X_{j'})|$. This setting isolates the effect of making relevance estimation more conservative, reducing the chance that a feature is selected due to an unstable correlation observed in a small source group.
- **UA-mRMR (UCB):** a variant in which the Fisher-z-based upper confidence bound is applied only to the per-group redundancy term, while relevance is computed using the standard $|\mathrm{corr}(X_j, y)|$. This setting isolates the effect of making redundancy estimation more conservative, penalizing features that may appear weakly redundant overall but are strongly redundant within particular source groups.
- **UA-mRMR (LCB+UCB):** the full design used in the main paper.

By comparing these variants, we can determine whether the performance gains arise primarily from a stricter relevance estimate, a stronger redundancy penalty, or the combination of both in the full uncertainty-aware objective.

Method	Logistic Regression		Random Forest	
	Accuracy	**AUROC**	**Accuracy**	**AUROC**
mRMR-Corr	0.737 ± 0.087	0.693 ± 0.110	0.744 ± 0.091	0.696 ± 0.121
UA-mRMR(LCB)	0.740 ± 0.087	0.698 ± 0.118	0.746 ± 0.090	0.686 ± 0.125
UA-mRMR(UCB)	0.739 ± 0.090	0.694 ± 0.117	0.745 ± 0.090	0.693 ± 0.121
UA-mRMR(LCB+UCB)	**0.756 ± 0.080**	**0.731 ± 0.093**	**0.758 ± 0.087**	**0.704 ± 0.117**

Table 4: Contribution of Relevance and Redundancy Adjustments in UA-mRMR on the TCGA-RNA modality. Values are mean ± standard deviation across 15 unseen target cancer types for three TCGA modalities.

As shown in Table 4, applying only one of the two Fisher-z adjustments yields limited benefit over the pooled mRMR-Corr baseline: UA-mRMR(LCB) and UA-mRMR(UCB) reach LR AUROC of 0.698 and 0.694, respectively, which are only marginally above mRMR-Corr's 0.693, and on RF AUROC both single-direction variants are actually slightly below the baseline. Combining the two directions, however, yields a substantial jump to LR AUROC 0.731 (+3.8% over mRMR-Corr) and RF AUROC 0.704 (+0.8% over mRMR-Corr). This pattern indicates that the two adjustments are synergistic: shrinking relevance estimates alone or inflating redundancy estimates alone removes only part of the unreliable signal, because the other term is still computed on pooled data that mixes heterogeneous source groups. Applying both adjustments simultaneously is necessary to realise the full transfer-performance gain, mirroring the decomposition result in Appendix B.1: group-aware averaging on both relevance and redundancy, together with the Fisher-z confidence bounds that shrink the benefit term and inflate the cost term, together produce the full UA-mRMR gain.

The relevance and redundancy adjustments above are applied at the bound width $\pm c\,\sigma_g$, with c=1 used throughout the paper. Because c controls how aggressively small-group estimates are discounted, we sweep it across all three TCGA modalities (5 seeds, 15 targets, LR and RF). UA-mRMR improves AUROC over mRMR-Corr for both classifiers across all mild widths ($c \leq 1$ SE), peaks at c=1 (Tables 5–6).

B.3. Statistical Significance of the Improvement over mRMR-Corr

Beyond mean-level comparisons, we test whether the improvement of UA-mRMR over mRMR-Corr is statistically significant across target cancer types. For each (downstream, metric) combination, the 15 paired target-level values are compared using a two-sided Wilcoxon signed-rank test. Table 7 reports the per-target mean and median differences, together with the Wilcoxon statistic W and the two-sided p-value.

UA-mRMR outperforms mRMR-Corr across target cancer types, with statistically significant improvements at the 5% level for both logistic-regression metrics (Wilcoxon

Method	LR AUROC	ΔLR	RF AUROC	ΔRF
mRMR-Corr (baseline)	0.673 ± 0.100	–	0.707 ± 0.104	–
UA-mRMR (c=0.5 SE)	0.685 ± 0.086	+0.012	0.708 ± 0.100	+0.001
UA-mRMR (c=1 SE, ours)	**0.692 ± 0.080**	**+0.019**	**0.712 ± 0.102**	**+0.004**
UA-mRMR (c=1.96 SE, 95%)	0.672 ± 0.093	-0.001	0.705 ± 0.100	-0.002
UA-mRMR (c=2.58 SE, 99%)	0.665 ± 0.090	-0.008	0.697 ± 0.102	-0.010

Table 5: Sensitivity to the confidence-bound width c for UA-mRMR. Values are mean ± standard deviation of AUROC across 15 unseen target cancer types for three TCGA modalities.

Method	LR AUPRC	ΔLR	RF AUPRC	ΔRF
mRMR-Corr (baseline)	0.658 ± 0.211	–	0.686 ± 0.203	–
UA-mRMR (c=0.5 SE)	0.671 ± 0.176	+0.013	**0.691 ± 0.177**	**+0.005**
UA-mRMR (c=1 SE, ours)	**0.674 ± 0.177**	**+0.016**	0.686 ± 0.175	+0.000
UA-mRMR (c=1.96 SE, 95%)	0.663 ± 0.179	+0.005	0.685 ± 0.173	-0.001
UA-mRMR (c=2.58 SE, 99%)	0.657 ± 0.174	-0.001	0.676 ± 0.172	-0.010

Table 6: Sensitivity to the confidence-bound width c for UA-mRMR. Values are mean ± standard deviation of AUPRC across 15 unseen target cancer types for three TCGA modalities.

Downstream	Metric	Δ(mean)	Δ(median)	W	raw p	Holm p	Bonf. p
LR	Accuracy	+0.019	+0.014	12.0	0.0043**	0.0171*	0.0171*
LR	AUROC	+0.038	+0.019	16.0	0.0103*	0.0308*	0.0410*
RF	Accuracy	+0.014	+0.014	27.0	0.0637	0.1274	0.2549
RF	AUROC	+0.008	+0.016	47.0	0.4887	0.4887	1.0000

Table 7: Paired Wilcoxon signed-rank test comparing UA-mRMR against mRMR-Corr across the 15 cancer datasets (TCGA-RNA). Δ = UA-mRMR − mRMR-Corr; Holm and Bonferroni columns adjust the raw p-values across the four tests

. Significance: $^{**}p < 0.01$, $^{*}p < 0.05$.

p=0.0043 for accuracy and p=0.0103 for AUROC). These survive multiple-testing correction: after Holm and Bonferroni adjustment across the four tests, both logistic-regression improvements remain significant at α=0.05 (Table 7). For random forest, the improvements remain positive in both metrics, but the magnitude is smaller and the paired test is underpowered at n=15 to detect the +0.008 AUROC gain as significant, consistent with tree ensembles' greater tolerance to feature redundancy.

B.4. Sensitivity to the Variant Pre-Filter Feature Size

The main experiments retain the top 500 variance features for the high-dimensional modalities (CNV, RNA). To confirm that the UA-mRMR vs. mRMR-Corr comparison is about the selection method rather than the variance filter, we re-ran both selectors across different feature size $\{100, 500, 1000\}$, all 3 modalities, K=25, 5 seeds, 15-cancer LOO. The UA-mRMR $\geq$ mRMR-Corr ordering is preserved at every feature size for average AUROC across three modalities (Table 8).

Cap	Per-modality ΔAUROC			Aggregate AUROC		
	CNV	**RNA**	**miRNA**	**UA-mRMR**	**mRMR-Corr**	**Δ**
100	+0.011	+0.006	-0.004	0.671	0.667	+0.004
500 (ours)	+0.003	-0.002	+0.009	0.694	0.690	+0.003
1000	+0.021	-0.002	+0.009	0.720	0.711	+0.009

Table 8: Sensitivity to Pre-filter feature size for UA-mRMR. Per-modality columns give ΔAUROC (UA-mRMR − mRMR-Corr, mean over LR+RF); the aggregate columns give AUROC averaged over the 3 modalities × LR+RF × 5 seeds × 15 targets. 500 is the setting we use in the main experiment.

Appendix C. Additional Prediction Target: Advanced Pathologic Stage

To assess whether these gains from UA-mRMR generalize beyond *TP53* mutation status, we evaluated a clinically distinct binary target, advanced pathologic stage (Stage III/IV), on the five cohorts with stage annotations (BLCA, COAD, ESCA, LUAD, SKCM), under the identical protocol (K=25, 5 seeds). On TCGA-CNV, UA-mRMR exceeds the baseline mRMR-Corr on this new target (Table 9). While target (whether it is Stage III/IV) is intrinsically harder to predict from features than mutation status, the UA-mRMR's advantage persists, demonstrating that our conclusions generalize beyond a single prediction target.

Clf	UA-mRMR AUROC	mRMR-Corr AUROC	Δ
LR	**0.522 ± 0.100**	0.517 ± 0.098	+0.005
RF	**0.525 ± 0.091**	0.520 ± 0.077	+0.005

Table 9: Performance comparison of UA-mRMR and mRMR-Corr while (Stage III/IV) as target.