

Biological Pathway Guided Gene Selection Through Collaborative Reinforcement Learning

Ehtesamul Azim
University of Central Florida
Orlando, Florida, USA
ehtesamul.azim@ucf.edu

Dongjie Wang
University of Kansas
Lawrence, Kansas, USA
wangdongjie@ku.edu

Tae Hyun Hwang
Vanderbilt University Medical Center
Nashville, Tennessee, USA
taehyun.hwang@vumc.org

Yanjie Fu
Arizona State University
Tempe, Arizona, USA
yanjie.fu@asu.edu

Wei Zhang[†]
University of Central Florida
Orlando, Florida, USA
wzhang.cs@ucf.edu

Abstract

Gene selection in high-dimensional genomic data is essential for understanding disease mechanisms and improving therapeutic outcomes. Traditional feature selection methods effectively identify predictive genes but often ignore complex biological pathways and regulatory networks, leading to unstable and biologically irrelevant signatures. Prior approaches, such as Lasso-based methods and statistical filtering, either focus solely on individual gene-outcome associations or fail to capture pathway-level interactions, presenting a key challenge: how to integrate biological pathway knowledge while maintaining statistical rigor in gene selection? To address this gap, we propose a novel two-stage framework that integrates statistical selection with biological pathway knowledge using multi-agent reinforcement learning (MARL). First, we introduce a pathway-guided pre-filtering strategy that leverages multiple statistical methods alongside KEGG pathway information for initial dimensionality reduction. Next, for refined selection, we model genes as collaborative agents in a MARL framework, where each agent optimizes both predictive power and biological relevance. Our framework incorporates pathway knowledge through Graph Neural Network-based state representations, a reward mechanism combining prediction performance with gene centrality and pathway coverage, and collaborative learning strategies using shared memory and a centralized critic component. Extensive experiments on multiple gene expression datasets demonstrate that our approach significantly improves both prediction accuracy and biological interpretability compared to traditional methods.

Release code and preprocessed data can be found at <https://github.com/ehtesam3154/bioMARL>

[†] Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '25, Toronto, ON, Canada

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1454-2/25/08
<https://doi.org/10.1145/3711896.3737198>

CCS Concepts

- Applied computing → Computational genomics; • Computing methodologies → Feature selection; Multi-agent reinforcement learning.

Keywords

automated gene selection; multi-agent reinforcement learning; biological pathway; disease outcome prediction

ACM Reference Format:

Ehtesamul Azim, Dongjie Wang, Tae Hyun Hwang, Yanjie Fu, and Wei Zhang[†]. 2025. Biological Pathway Guided Gene Selection Through Collaborative Reinforcement Learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3711896.3737198>

1 Introduction

With the rapid advancement of high-throughput technologies, high-dimension, low-sample-size (HDLSS) data have become increasingly prevalent in biomedical research. Large cohort cancer studies, such as The Cancer Genome Atlas (TCGA) [10], contain vast amounts of HDLSS genomic data, offering an unprecedented opportunity to understand cancer mechanisms and improve therapeutic outcomes [1, 7, 21]. However, extracting meaningful insights from this large genomic data requires addressing major analytical challenges, such as statistical randomness, experimental noise, and sample heterogeneity, which often result in inconsistent and biologically or clinically irrelevant gene signatures [53].

A well-known example of this challenge is the discrepancy between two early breast cancer diagnostic gene panels: MammaPrint, a 70-gene signature developed in the Netherlands [42], and a 76-gene signature from a similar study in San Diego [45]. Despite their shared goal of predicting disease progression, only three genes overlapped, highlighting the difficulty of selecting stable gene signatures (i.e., biomarkers). Traditional feature selection methods, such as Lasso-based approaches [28, 41], have been widely used for biomarker selection due to their ability to enforce sparsity and identify predictive genes. However, these methods primarily focus on individual gene-outcome associations while neglecting the complex biological pathways and regulatory networks that govern gene function. Genes do not operate in isolation; rather, they interact within intricate systems, and integrating biological pathway information

as prior knowledge can enhance the robustness and interpretability of selected gene signatures [13]. By leveraging known gene interactions and functional relationships, gene selection methods can improve reproducibility and biological relevance.

That said, integrating biological pathway knowledge into gene selection presents unique challenges that extend beyond simple statistical incorporation. While databases like KEGG [24] provide extensive biological pathway information, effectively leveraging this knowledge requires understanding complex dependencies between genes and their collective impact on biological processes. Most traditional feature selection methods operate independently of biological pathway knowledge, relying purely on statistical patterns in the data [52]. Simple filtering or scoring mechanisms that attempt to integrate pathway information often fail to capture the dynamic nature of gene interactions, leading to suboptimal gene selection. This highlights the need for a framework that models these complex dependencies while making biologically informed decisions about gene importance, ensuring selected gene signatures reflect mechanisms rather than purely statistical associations.

Our contribution: A pathway-aware gene selection perspective. We approach the gene selection problem from the perspective of collaborative decision-making, where each gene acts as an intelligent agent that learns when its inclusion benefits both predictive power and biological relevance. The key insight is that: genes function through complex regulatory pathways and networks, interacting via molecular mechanisms, protein-protein interactions, and signaling cascades to drive cellular functions. Our selection agents mirror this biological reality by learning to make decisions that consider both their individual contributions and their collective impact within these pathways. By prioritizing genes with established pathway annotations, our framework leverages decades of experimental validation and enables more interpretable results, as selected genes can be understood within known biological processes. This biological context-aware selection also facilitates cross-study validation and increases the likelihood of identifying druggable targets, as existing therapeutics often target specific pathways.

This paper presents a two-stage architecture where pathway-guided statistical pre-filtering creates a biologically relevant reduced gene set, followed by a MARL-based selector where genes function as collaborative agents through sophisticated mechanisms including intelligent reward design, shared memory, and centralized feedback to maintain both statistical significance and biological relevance while optimizing predictive performance. Our key contributions are as follows:

- First framework to unify statistical feature selection with pathway knowledge in genomic data analysis.
- Novel application of MARL to model gene selection as a collaborative decision process.
- Introduction of biologically-informed learning mechanisms that maintain both predictive power and pathway relevance.

2 Problem Statement

This paper addresses the challenge of selecting relevant genes from high-dimensional genomic datasets. The goal is to identify a subset of genes that optimize predictive performance while maintaining biological interpretability and pathway-level coherence, addressing

the limitations of traditional feature selection methods in capturing complex gene interactions. Formally, given a high-dimensional genomic dataset D with gene set G , where each sample consists of gene expression values and an associated outcome, our objective is to select an optimal subset $G_{opt} \subset G$ such that $|G_{opt}| \ll |G|$, significantly reducing the dimensionality while preserving predictive power. Through a two-stage selection process, we first obtain a pre-filtered set $G_{pre} \subset G$ using pathway-guided statistical methods, followed by a refined selection process that produces the final ranked set G_{ranked} from which the top k genes form G_{opt} . The selected genes should not only optimize the performance of downstream prediction tasks but also reflect meaningful biological pathway relationships, ensuring both statistical significance and biological relevance in the final gene signature.

3 Methodology

3.1 Framework Overview

Figure 1 illustrates our pathway-aware gene selection framework **BioMARL** consisting of two main components: The first component implements a biological pathway-guided gene pre-filtering that combines multiple statistical selection methods with pathway information from the KEGG database. This initial filtering creates a biologically relevant reduced gene set by evaluating both statistical significance and pathway-level performance. The filtered gene set then serves as input to our MARL based selector, where each gene is modeled as an independent agent within a collaborative framework. This incorporates pathway knowledge at multiple levels: through pathway-aware state representations using Graph Neural Networks and a reward mechanism that considers both statistical and pathway-based performance. The framework is further enhanced with collaborative learning mechanisms including a shared memory system for effective knowledge sharing between agents and a centralized critic for evaluating collective agent behavior. This ultimately outputs a ranked list of selected genes that demonstrate both strong predictive power and biological relevance.

3.2 Pathway-Guided Gene Pre-filtering

Why integrate biological context with statistical selection? Purely statistical feature selection in genomic data analysis faces two fundamental challenges. First, the curse of dimensionality in HDLSS genomic datasets results in unstable feature¹ rankings, where small variations in data can lead to substantially different selected gene sets. Second, complex non-linear interactions between genes make it difficult to accurately determine gene importance using individual statistical measures alone. Although advanced statistical techniques, such as regularization and ensemble methods, attempt to mitigate these issues, they remain biology-agnostic and fail to capture functional gene relationships. A biologically informed approach, leveraging pathway knowledge, introduces natural constraints and grouping structures to address these challenges. It reduces the effective search space by prioritizing biologically plausible feature combinations, while the inherent pathway-based organization of genes facilitates the identification of functionally

¹In this work, we use ‘feature’ and ‘gene’ interchangeably, as ‘feature’ is the standard term in the broader feature selection literature, including gene selection studies.

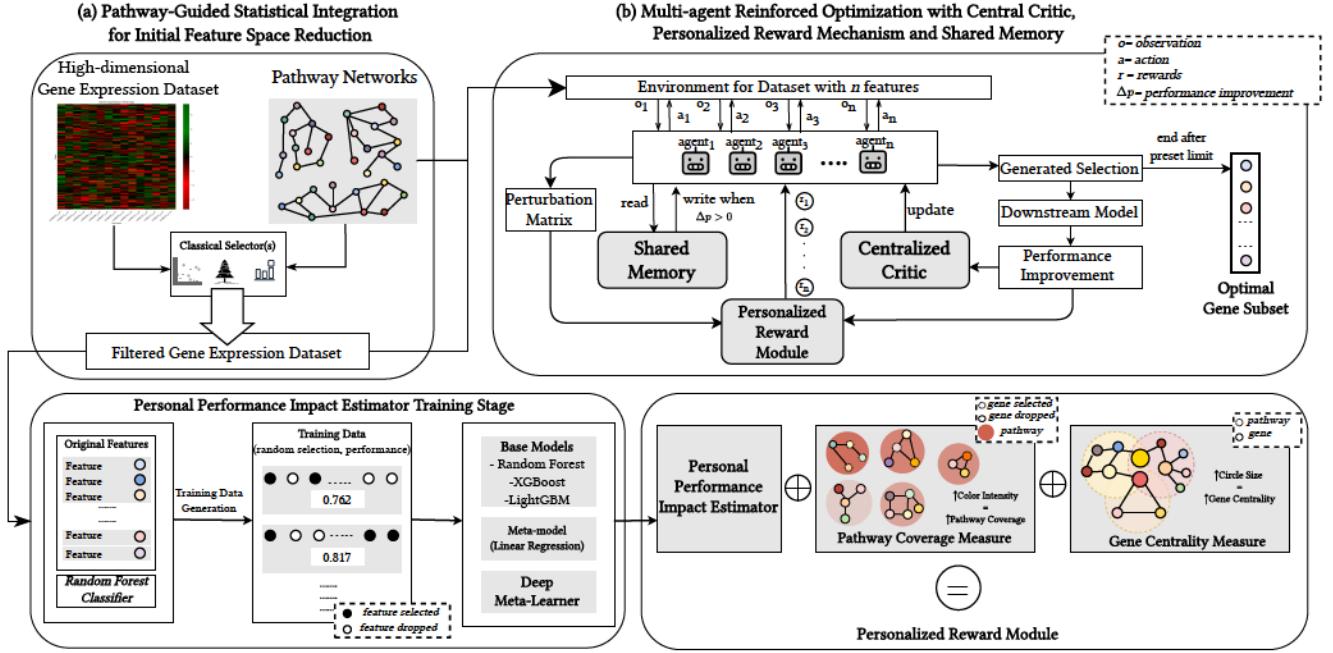


Figure 1: Framework Overview. BioMARL consists of two key parts: (a) pathway-guided meta-selection combining multiple statistical methods with KEGG pathway information for initial filtering; (b) a pathway-aware multi-agent RL framework for refined selection, with centralized evaluation, shared memory and a multi-component reward system to enhance collaboration among agents while preserving biological relevance. The iteratively generated feature set is evaluated in downstream task, with iterations continuing until optimization or set limit.

relevant interactions that purely statistical approaches may overlook. By embedding pathway information, gene selection becomes more stable, interpretable, and computationally efficient, focusing on gene sets that align with known biological mechanisms rather than purely data-driven statistical associations.

We implement a pathway-guided pre-filtering strategy that combines statistical feature selection with biological pathway knowledge. The pre-filtering process consists of three main components:

Base Score Computation. Multiple statistical feature selection methods are employed to compute initial importance scores for each gene. For a dataset D with gene set G , each method $f_i \in F$ generates a score vector S_i , where $S_{i,g}$ represents the importance score assigned to gene g by method f_i . Our framework employs three methods ($|F| = 3$): chi-squared test, random forest importance, and SVM-based feature ranking.

Pathway Performance Integration. We leverage the KEGG pathway database $P = \{p_1, \dots, p_m\}$ to incorporate biological context. For each pathway p_i containing a gene set G_i , we compute a score $S_p(p_i)$ using a classifier trained on genes within that pathway: $S_p(p_i) = \text{classifier_performance}(D[G_i])$, where $D[G_i]$ represents the dataset restricted to genes in pathway p_i .

Integrative Score Calculation. The final importance score \hat{s}_g for each gene g combines both statistical and pathway-based evidence. Each method's weight is computed based on its performance: $w_i = \frac{S_p(f_i)}{\sum_{f_j \in F} S_p(f_j)}$, where $S_p(f_i)$ is the performance score of method f_i on validation data. The base meta-score for gene g is: $m_g = \sum_i w_i \cdot S_{i,g}$. This score is then adjusted using pathway information: $\hat{s}_g = m_g \cdot (1 + \beta \cdot \log(1 + \bar{S}_p(g)))$, where $\bar{S}_p(g)$ is the mean performance score

of pathways containing gene g , and β is a scaling factor. Unmapped genes retain their base meta-scores to avoid penalizing potentially important but less studied genes. The final filtered gene set is: $G_{pre} = \{g \in G : \hat{s}_g > \mu + 2\sigma\}$ where μ and σ are the mean and standard deviation of the adjusted scores.

3.3 Multi-Agent Gene Selection Framework

Why model genes as collaborative decision-makers? Building on the pre-filtered gene set G_{pre} , we implement a MARL framework where each gene in G_{pre} acts as an agent that learns to make selection decisions collaboratively. Traditional feature selection methods treat genes independently, failing to capture the complex interactions within biological pathways. MARL allows us to model genes as collaborative agents to reflect both their individual significance and collective behavior in pathways. Through pathway-aware state representations and a reward mechanism that integrates both statistical and pathway-level performance, agents learn selection strategies that balance predictive power with biological relevance while efficiently handling high-dimensional genomic data.

Action Space. The action space for each agent i (i.e., gene i) is binary: $a_i \in \{0, 1\}$, where 0 represents discarding and 1 represents selecting the gene.

State Representation. To effectively capture both gene interactions and pathway relationships, we employ a Graph Neural Network (GNN) for state representation. The state s_t at time step t is constructed as: $s_t = \text{GNN}(X_t, E_t)$, where X_t represents the expression matrix of genes in G_{pre} augmented with pathway embeddings at time step t . The edge connections E_t are constructed through an

adaptive mechanism that combines correlation and pathway information: $E_{ij} = \rho \cdot C_{ij} + (1 - \rho) \cdot J_{ij}$ where C_{ij} is the correlation between genes i and j , J_{ij} is the Jaccard similarity of their pathway memberships, and ρ is the correlation weight. This information is then processed through graph convolutional layers:

$$h = \sigma(\tilde{A} \times W^{(0)})$$

$$s_t = \text{pool}(\sigma(\tilde{A}hW^{(1)}))$$

where \tilde{A} is the normalized adjacency matrix derived from the edge connections, $W^{(0)}$ and $W^{(1)}$ are learnable weight matrices, and σ is the ReLU activation function. The pooling operation aggregates node-level features into a global state while preserving both individual gene characteristics and pathway contexts.

Multi-Agent Deep Q-Network Architecture. Each agent employs a Deep Q-Network (DQN) to learn optimal selection strategies. The Q -function for agent i (i.e., gene i) is approximated by a neural network $Q_i(s, a; \theta_i)$ that maps states to action values, where θ_i are the network parameters. The network learns to estimate the expected cumulative reward for taking action a in state s :

$$Q_i(s, a; \theta_i) \approx \mathbb{E}[R_t \mid s_t = s, a_t = a]$$

where $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$ is the discounted cumulative reward. Drastic reduction in feature space while trying to maintain or improve prediction performance does not exactly go hand-in-hand. To stabilize learning and improve efficiency, we implement prioritized experience replay and maintain a target network. Prioritized replay ensures that important transitions are sampled more frequently during training, while the target network, updated periodically, reduces overestimation bias in the Q -value updates. The networks are trained using the Huber loss to provide robustness against outliers, and an epsilon-greedy policy with decaying exploration rate ensures sufficient exploration of the feature space while gradually focusing on exploiting learned strategies.

Centralized Critic. To provide a global perspective on the collective performance of our multi-agent feature selection system, we implement a centralized critic $V(s)$ that maps the current state s_t to a scalar value estimate v_t . The critic architecture combines compression, gating, and dynamic layers to effectively process the high-dimensional state information. It is trained to minimize the mean squared error between its prediction and observed performance improvement: $I_{\text{critic}} = \text{MSE}(v_t, I_t)$, where I_t is global performance improvement at time step t .

The critic's value estimates are incorporated into the DQN learning process, influencing the target Q -values: $Q_{\text{target}} = \lambda_a(r_t + \gamma \max_a Q(st+1, a)) + \lambda_b v_t$, where λ_a and λ_b are weighting factors. This integration helps balance local and global optimization, promoting more stable and effective feature selection by providing a baseline that reduces variance in the policy updates. The global value estimate provided by the critic serves as a form of baseline, helping to reduce variance in the policy gradients implicitly computed by the DQNs through their value function updates.

Shared Memory. To enhance collaboration between agents and leverage collective knowledge, we introduce a shared memory mechanism consisting of two components: a collaboration success record and a synergy matrix. The collaboration success record H is defined as a mapping from gene sets to their corresponding performance improvements: $H : \mathcal{P}(F) \rightarrow \mathbb{R}$ where F is the set of

all features, $\mathcal{P}(F)$ power set of F , and \mathbb{R} represents performance improvements. For a given feature set $S \in \mathcal{P}(F)$, the collaboration success is updated as $H(S) = \max(H(S), \Delta p)$, where Δp is the observed performance improvement when using feature set S . The synergy matrix $M \in \mathbb{R}^{d \times d}$, where d is the number of features, captures pairwise feature interactions. For features i and j , the synergy value updated as: $M[i, j] += \Delta p$, when features i and j are both present in a successful feature combination. Both components follow an exponential decay mechanism to adapt to changing feature relationships over time.

The shared memory influences the action selection process by introducing a synergy bias. The probability of selection action a for feature i is given by: $P(a_i = 1) = \sigma(Q(s_i, 1) + \eta \cdot \sum_{j \in P_i} M[i, j])$, where $Q(s_i, 1)$ is the Q -value for selecting feature i , P_i refers to the top k partners to consider for feature i based on the synergy matrix, σ is the sigmoid function, and η is a hyperparameter controlling the influence of the synergy bias.

Reward Mechanism. To address the credit assignment problem in MARL-based feature selection, we implement a sophisticated reward mechanism that combines efficient performance estimation with pathway-level insights to guide agent decisions. The reward calculation consists of three key components:

Personal Performance Impact Estimator. First, to estimate individual feature contributions efficiently, we construct a perturbation matrix D_t , where each row flips a single element of the selection vector $a_t \in \{0, 1\}^d$, representing an alternative feature state. Evaluating model performance for each state yields labels y_t , and (D_t, y_t) is passed as training data to an ensemble meta-learner f_{meta} (combining random forest, XGBoost, LightGBM with a meta model and a neural network). Performance changes are estimated as $\Delta R_t = f_{\text{meta}}(D_t) - R_t$, where R_t is the current performance. By leveraging matrix-based updates and parallel computation, our approach avoids the inefficiencies of sequential feature evaluations, making it scalable for high-dimensional genomic data.

The reward for each agent incorporates uncertainty-aware adjustments. Given an uncertainty vector u_t estimated by f_{meta} , we compute the confidence factor $c_t = \frac{1}{1+u_t}$ and define the reward as: $r_{\text{base}} = \Delta R_t \odot c_t - \log(1+u_t) + I_t 1$, where I_t represents global performance improvement.

To maintain accuracy throughout the selection process, we implement an online learning mechanism. Periodically, at a pre-defined update frequency F , we update the meta-learner using a buffer $\mathcal{B} = \{(a_t, R_t)\}_{t=1}^B$ of recent experiences, where a_t represents the joint action vector of all agents at time t , and B is the buffer size. The update minimizes the loss:

$$L = \sum_{(a_t, R_t) \in \mathcal{B}} (f_{\text{meta}}(a_t) - R_t)^2 + \lambda \Omega(f_{\text{meta}})$$

where $\Omega(f_{\text{meta}})$ is a regularization term with coefficient λ .

Gene Centrality Measure. We introduce a pathway-based centrality measure that evaluates each gene's importance within and across biological pathways. Unlike traditional centrality metrics that focus solely on network topology, our measure considers both the gene's pathway membership and its connectivity patterns to capture its functional significance. For a gene i , we define a centrality score ϕ_i that captures its importance within and across pathways: $\phi_i = n_i \times$

$\sum_{p \in P_i} |G_{i,p}|$, where n_i is the number of pathways containing gene i , P_i is the set of pathways containing gene i , and $|G_{i,p}|$ represents the number of genes connected to gene i in pathway p .

For a selection S , the aggregate centrality function $\Phi(S)$ incorporates both pathway membership and cross-pathway connectivity:

$$\Phi(S) = \sum_{i \in S} \left(n_i \sum_{p \in P_i} |G_{i,p} \cap S| \right)$$

For each gene i , we calculate a differential centrality score $\Delta\phi_i$ that quantifies how its inclusion or exclusion impacts the overall pathway representation. For unselected genes, we calculate:

$$\Delta\phi_i = \Phi(S \cup \{i\}) - \Phi(S)$$

where $\Phi(S)$ represents the aggregate centrality score for the current selection S . For genes already in the selection, we calculate the loss in centrality from their removal:

$$\Delta\phi_i = \Phi(S) - \Phi(S \setminus \{i\})$$

This formulation rewards genes that serve as critical connectors between multiple pathways, reflecting their potential regulatory importance in biological processes.

Pathway Coverage Measure. We implement a pathway coverage measure to ensure comprehensive representation of biological processes by evaluating how effectively the selected genes span known biological pathways. For a given pathway p , we define a coverage score ψ_p that measures the proportion of pathway genes included in the current selection: $\psi_p(S) = \frac{|S \cap G_p|}{|G_p|}$, where G_p represents the set of genes in pathway p , and S is the current gene selection. For each gene i , we calculate a differential coverage score $\Delta\psi_i$ that quantifies its contribution to pathway coverage. For unselected genes, we calculate the potential coverage gain:

$$\Delta\psi_i = \psi_p(S \cup \{i\}) - \psi_p(S)$$

For genes already in the selection, we calculate the coverage loss from their removal:

$$\Delta\psi_i = \psi_p(S) - \psi_p(S \setminus \{i\})$$

This measure encourages the selection of genes that maximize the representation of diverse biological processes while penalizing redundant selections within already well-covered pathways.

The final reward for each gene i combines the base performance estimate with the pathway-based measures through a weighted sum: $r_i = \omega \cdot r_{base} + \xi \cdot \Delta\phi_i + \zeta \cdot \Delta\psi_i$, where ω, ξ, ζ are weighting coefficients that sum to 1, balancing the importance of statistical performance (r_{base}), gene centrality impact ($\Delta\phi_i$), and pathway coverage contribution ($\Delta\psi_i$).

Optimal Set Selection. The final optimal gene set is determined by computing weighted importance scores for each gene based on the Q -value differences between its selection and rejection actions over time: $w_i = \sum_{t=1}^T \gamma^{T-t} (Q_i^t(s_t, 1) - Q_i^t(s_t, 0))$, where γ is a decay factor giving more weight to recent decisions. Genes are ranked by these scores and the top k genes form the optimal set $G_{opt} = \{g_i \in G_{ranked} : i \leq k\}$ are selected. This captures the cumulative learning of agents while favoring consistently important genes.

4 Experiment

4.1 Experimental Setup.

Dataset Descriptions. We performed experiments on the TCGA breast invasive carcinoma (BRCA) [9], lung adenocarcinoma (LUAD) [11] and ovarian serous cystadenocarcinoma (OV) [12] datasets. For the BRCA studies, patients were classified based on estrogen receptor (ER+ versus ER-), progesterone receptor (PR+ versus PR-), human epidermal growth factor receptor 2 (HER2+ versus HER2-), and triple negative (TN versus non-TN) status. For the LUAD and OV studies, we stratified patients into two groups based on survival time: those who survived less than 25 months and those who survived more than 50 months. Each dataset has 20,530 features (i.e., genes). More detailed statistics shown in Table 1.

Table 1: Sample distribution across datasets. Class 1: ER+, HER2+, PR+, TN, survival time <25 months. Class 2: ER-, HER2-, PR-, non-TN, survival time >50 months.

Dataset	Class 1	Class 2
BRCA (ER)	332	80
BRCA (HER2)	334	78
BRCA (PR)	285	127
BRCA (TN)	65	347
LUAD	109	21
OV	63	49

Baseline Algorithms. We compared BioMARL with eight widely used feature selection methods: 1. **K-Best** [48] selects K features with the highest feature score; 2. **mRMR** [35] selects a feature subset with the highest relevance to the target and least redundancy among themselves; 3. **LASSO** [41] uses regularization to shrink coefficients of less useful features to zero, effectively performing feature selection during model fitting; 4. **RFE** [20] recursively removes the weakest features until a specified number of features is reached; 5. **LASSONet** [28] is a neural network with sparsity to encourage the network to use only a subset of input features; 6. **GFS** [5] selects features using genetic algorithms, which recursively generates a population based on a possible feature subset, then uses a predictive model to evaluate it; 7. **RRA** [38] collects distinct selected feature subsets, and then integrates them based on statistical sorting distribution; 8. **MCDM** [22] approaches feature selection as a Multi-Criteria Decision-Making problem and uses the VIKOR sort algorithm to rank features based on the judgment of multiple selection methods. All experiments use AUC score as the metric to evaluate the classification performance. The baselines along with our method, were executed 10 times on each dataset for every experiment, and the average AUC score across these runs was reported for each method. We adopted Random Forest as the downstream classification model and during each run, a 70-30 training-testing split was employed, with genomic feature selection performed using 5-fold cross-validation on the training set, followed by evaluation of the selected features on the hold-out set. All reported performance scores correspond to $k = 100$ selected features.

4.2 Experimental Results

Overall Comparison. In this section, we assess the performance of BioMARL and baseline algorithms for feature selection on the

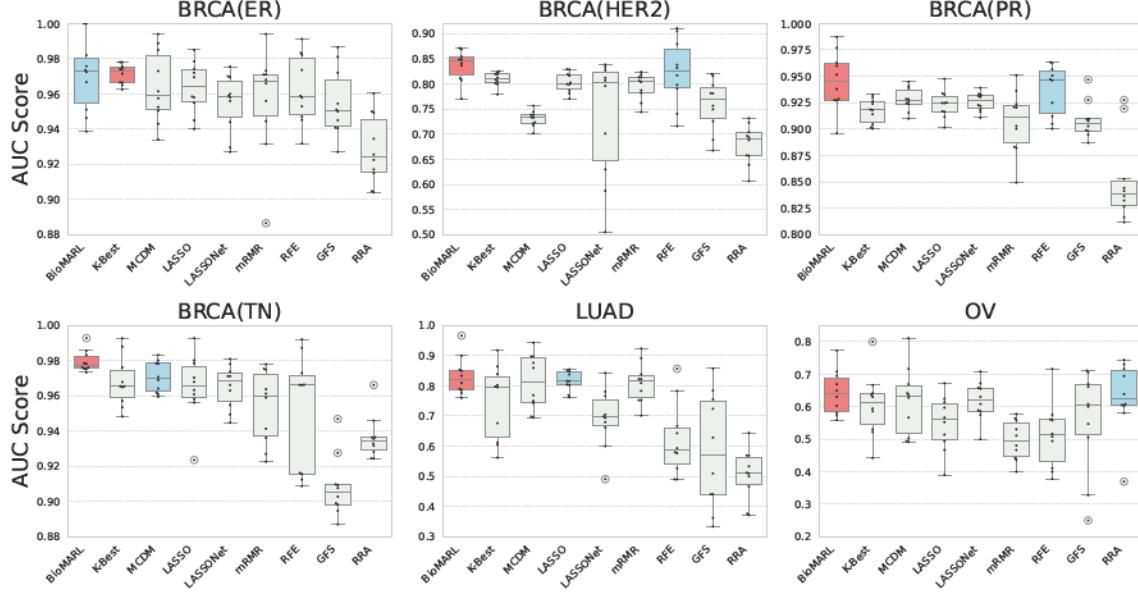


Figure 2: Overall performance comparison of BioMARL with eight state-of-the-art baselines: Best performing baseline highlighted in **lightcoral** and second-best in **lightblue**.

aforementioned datasets. Figure 2 shows the comparison result. BioMARL outperforms other selection algorithms in 5 out of the 6 datasets, demonstrating consistently higher AUC scores. Additionally, it exhibits relatively low variance in most cases, indicating stable and reliable performance. The underlying driver behind this performance is BioMARL’s dual-layer optimization approach. The pathway-guided pre-filtering first creates a high-quality candidate pool by eliminating statistically insignificant genes while preserving meaningful pathway relationships. This provides a strong foundation for feature selection which the subsequent MARL framework then leverages through sophisticated reward estimation that considers both individual gene contributions and their synergistic effects, allowing agents to identify feature combinations that maximize classification performance.

Ablation Study. To validate the impact of each technical component, we developed three variants of BioMARL: (i) **BioMARL**–*Rwd* uses performance improvement directly as the reward for RL agents, (ii) **BioMARL**–*Crt* operates without a centralized critic, and (iii) **BioMARL**–*Mem* excludes the shared memory mechanism. Table 2 presents the comparison results for two breast cancer studies, as well as the LUAD and OV datasets. BioMARL outperforms **BioMARL**–*Rwd*, demonstrating that the pathway-aware reward mechanism effectively balances statistical significance with biological relevance, resulting in more robust and biologically meaningful feature selection. The superior performance over **BioMARL**–*Crt* highlights the importance of centralized critic in capturing global state information, resulting in better-coordinated agent behaviors and avoiding conflicting local optimizations. Finally, advantage over **BioMARL**–*Mem* demonstrates the value of shared memory mechanism, which enables collaborative learning and accelerates the discovery of synergistic feature sets by avoiding redundant exploration of suboptimal combinations.

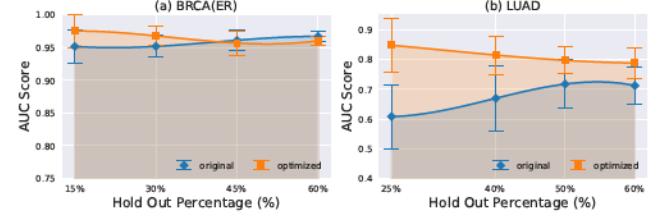


Figure 3: The influence of different hold-out percentages in terms of AUC score. ‘optimized’ refers to the subset of features selected by BioMARL.

Study of the Impact of Hold-out Percentage. In gene selection problems with imbalanced datasets and limited samples, evaluating feature selection algorithms across various train-test splits, including unconventional ratios, might be beneficial. This highlights the model’s robustness in scenarios with smaller sample sizes, where traditional splits (80-20 or 70-30) may not suffice. Additionally, it offers practical insights into model performance when data availability is limited, which is common in expensive and data-scarce applications. In the experiment setup section, we used a 70-30 train-test split. We further explored the impact of the holdout setting through experiments on BRCA(ER) and LUAD, as shown in Figure 3. While they exhibit varying patterns due to factors like randomness of dataset partitioning, training data volume, and task complexity, BioMARL consistently demonstrates performance enhancements. Despite the lack of uniform trend across them, the results generally indicate that BioMARL reliably improves performance across most scenarios. This consistency affirms the robustness of our approach regardless of the holdout percentage used.

Clear Separation of Breast Cancer Patients by BioMARL-Selected Genes. The heatmap of gene expression data for top 100 genes selected by BioMARL exhibits a well-defined pattern, achieving perfect separation between PR+ and PR- breast cancer patients.

Table 2: Ablation study of BioMARL. Best performance highlighted in bold

Variant	Technical Component			Performance			
	Personalized Reward	Centralized Critic	Shared Memory	BRCA(ER)	BRCA(HER2)	LUAD	OV
BioMARL ^{-Rwd}	✗	✓	✓	0.9598 ± 0.023	0.8184 ± 0.061	0.7064 ± 0.107	0.6276 ± 0.106
BioMARL ^{-Crt}	✓	✗	✓	0.9605 ± 0.029	0.8073 ± 0.030	0.7623 ± 0.058	0.6298 ± 0.095
BioMARL ^{-Mem}	✓	✓	✗	0.9618 ± 0.0166	0.7914 ± 0.093	0.7712 ± 0.089	0.6494 ± 0.100
BioMARL	✓	✓	✓	0.9706 ± 0.0198	0.8357 ± 0.0296	0.8327 ± 0.0601	0.6439 ± 0.065

Most of the selected genes show elevated expression in PR- patients, suggesting their involvement in pathways related to PR signaling and breast cancer progression. This clear distinction underscores BioMARL’s effectiveness in identifying biologically meaningful and clinically relevant features, reinforcing its potential as a powerful tool for biomarker discovery. Identifying genes that distinguish PR subtypes is critical for understanding the molecular mechanisms driving breast cancer heterogeneity and could provide valuable insights for prognosis and targeted therapy development. Notably, similar clear expression patterns were also observed in other breast cancer subtypes which further validate BioMARL’s robustness in capturing key regulatory signatures, highlighting its significance in advancing precision oncology and personalized treatment. A potential reason behind this is its collaborative multi-agent architecture that enables agents to identify complementary genes that collectively maximize class separation, rather than relying on individual significance, forming robust discriminative patterns.

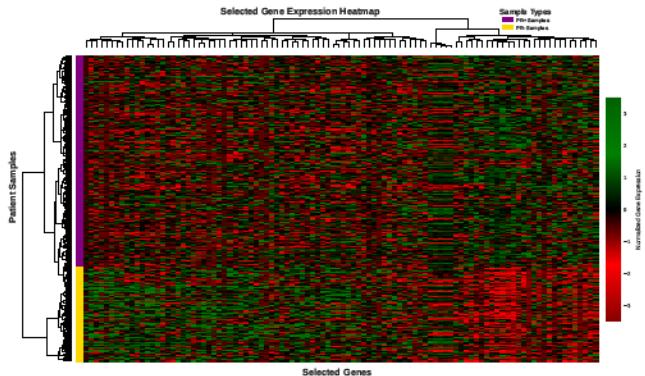
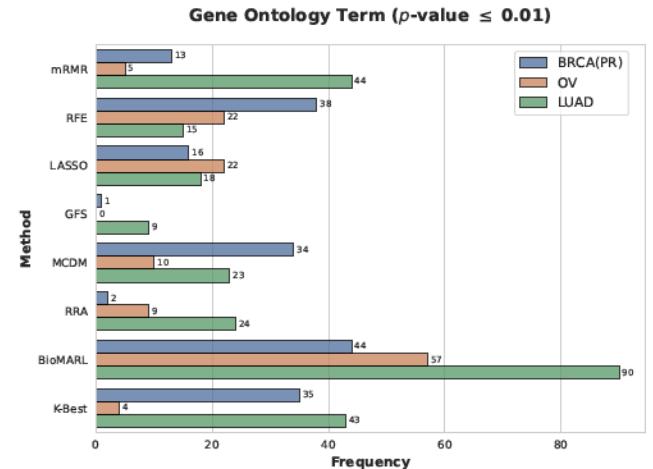


Figure 4: Heatmap of the expression profiles of 100 marker genes selected by BioMARL on the BRCA(PR) dataset.

Gene Enrichment Analysis. To evaluate the biological relevance of genes selected by BioMARL and baseline methods, we performed gene set enrichment analysis. This analysis aims to identify significant overlaps between selected genes and known functional gene sets, which may provide insights into the biological processes and pathways potentially involved in these cancers. We examined the enrichment of Gene Ontology (GO) terms [3] among the top-ranked 100 genes identified by each method. Figure 5 illustrates the frequency of enriched GO terms $p\text{-value} \leq 0.01$ for all different methods across three datasets. BioMARL demonstrated superior performance on both datasets which suggests that it effectively captures genes involved in biologically relevant processes specific to different cancer types. The underlying impetus for this is BioMARL’s pathway-guided pre-filtering that establishes a biologically relevant initial feature space, and its ability to combine biological domain

knowledge with collaborative learning, where intelligent agents make selection decisions guided by both pathway information and collective performance optimization.

Figure 5: Enrichment analysis of selected genes for breast, ovarian and lung cancer datasets. Bar plots show the frequency of significantly enriched Gene Ontology terms ($p\text{-value} \leq 0.01$) for different selection methods.

Robustness Check of BioMARL Under Different Classification Models. Feature selection methods depend on evaluating the currently selected feature subset at each step with a downstream predictive model to identify the optimal feature subset. In our design, we employed the Random Forest for evaluation due to its stability and robustness. However, more rigorously, a feature selection method should be able to generalize to diverse downstream models. We examine the robustness of BioMARL on BRCA(TN) by changing the classifier to XGBoost, SVM, Decision Tree, LASSO and Ridge Regression. The comparison results, depicted in Figure 6 shows that BioMARL exhibits downstream model-agnostic robustness and reliability. A potential reason behind this robustness is our multi-agent framework’s ability to learn feature importance through collaborative decision-making and shared memory, rather than being tied to any specific model’s optimization criteria. The agents learn selection strategies that identify fundamentally predictive features through their collective experience, leading to stable performance regardless of the downstream classifier used.

Case Study I: BioMARL Selected Genes Have Proven Relevance in Cancer. To further validate the biological relevance of genes uniquely selected by BioMARL, we conducted a literature review focusing on these genes in the context of breast cancer. Table 3 highlights a subset of the top-ranked genes and their known associations with breast cancer pathogenesis. Our analysis revealed

Table 3: Literature review of the candidate cancer genes. This table presents citations highlighting the relevance of the genes selected exclusively by BioMARL.

Gene Name	Pathway	Description	Reference
<i>CYP3A7</i>	hsa00140	When expressed in adults, driven by the <i>CYP3A7*1C</i> allele, is associated with altered hormone metabolism and impacts breast cancer outcomes.	[23]
<i>NME1</i>	hsa00230	Upregulated in DCIS but downregulated in invasive breast cancer; loss of <i>NME1</i> promotes tumor invasion by increasing <i>MT1-MMP</i> surface levels.	[32]
<i>FGA</i>	hsa04610	Identified as a key biomarker in HER2-positive breast cancer plasma samples, with lower levels in cancer patients compared to controls, reverting to control levels post-surgery.	[43]
<i>GRB7</i>	hsa04140	Co-amplified with HER2 gene in breast cancer, <i>GRB7</i> facilitates HER2-mediated signaling and tumor formation, with its protein overexpression concurrent with HER2 amplification.	[8]
<i>ERBB2</i>	hsa04520	Major oncogene amplified in 20–30% of breast cancers, its overexpression correlates with tumor chemoresistance and poor patient prognosis, making it a crucial therapeutic target.	[15]
<i>AGTR1</i>	hsa04614	Shows profound overexpression in a subset of breast tumors across independent cohorts, suggesting a potential role in tumorigenesis and as a therapeutic target.	[37]
<i>PSMC1</i>	hsa03050	Significantly upregulated in breast cancer tissues, with high expression correlating with poor survival outcomes and involvement in critical cancer-related cellular processes.	[25]
<i>TUBB3</i>	hsa04540	High <i>TUBB3</i> expression in breast cancer correlates with decreased sensitivity to taxane-based chemotherapy and is associated with high tumor grade and advanced tumor stage.	[27]

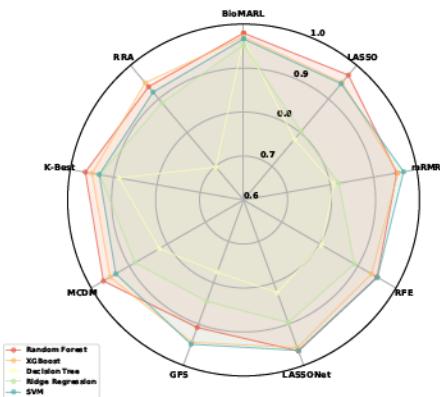


Figure 6: Robustness evaluation of BioMARL across classification models, measured by AUC score on BRCA(TN). Larger area coverage indicates better overall performance across feature selection methods.

that several of these genes have established roles in breast cancer progression, metastasis, or treatment response. For instance, the adult expression of *CYP3A7* is linked to altered estrogen metabolism, potentially influencing breast cancer outcomes and response to specific chemotherapeutic agents [23]. *ERBB2* is amplified in 20–30% of breast cancers, driving chemoresistance and poor prognosis, making its overexpression a key therapeutic target [15]. *GRB7* not only facilitates HER2-mediated signaling and tumor formation in breast cancer cells but can also be selectively retained and overexpressed in some solid tumors independent of HER2 amplification [8]. Additionally, *PSMC1* exhibits significantly higher expression in breast cancer tissues compared to normal tissues, with high transcript levels correlating with poor survival outcomes, suggesting its potential as a prognostic biomarker for aggressive disease [25].

Case Study II: BioMARL Improved Survival Prediction. To evaluate the prognostic significance of the top genes selected by BioMARL, we conducted a survival analysis on cancer datasets. Patients' overall survival was assessed using Kaplan-Meier (KM)

plots, stratified by high and low expression levels of the top-ranked genes identified by BioMARL. As shown in Figure 7, KM survival curves were generated for both groups, and the log-rank test was performed to compare their survival distributions. The KM curves revealed a clear separation between the two groups for the selected genes, indicating potential prognostic value. Additionally, the log-rank test produced significant *p*-values for these genes, reinforcing their association with overall survival in different cancer patients. This analysis demonstrates that genes selected by BioMARL are not only relevant to the disease but also have significant implications for patient outcomes.

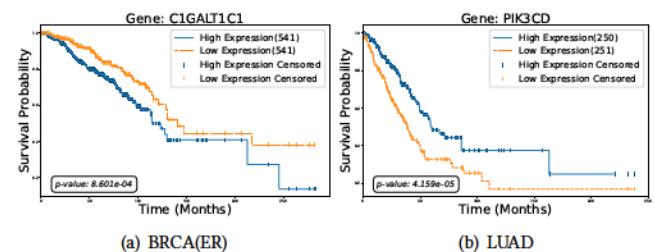


Figure 7: Survival analysis on (a) breast cancer and (b) lung cancer patients with gene selected by BioMARL

Real-World Deployment and Validation of BioMARL in Gastric Cancer Patient Data. Validating a single gene as a cancer biomarker requires substantial financial investment, as biologists and biomedical scientists must conduct gene knockout or knockdown experiments to assess its effects on cell viability, proliferation, and apoptosis *in vitro*. These studies are typically performed in mouse models or cancer cell lines, rather than in actual cancer patients. Identifying a viable drug target and developing a therapeutic for real patients requires clinical trials, which are essential for evaluating safety and efficacy. Over the past two decades, the average cost of drug development has ranged from \$539 million to \$2.8 billion [33]. The proposed BioMARL can help prioritize target

genes, significantly reducing wet-lab experiment costs and lowering research and development (R&D) expenses in drug discovery.

To evaluate the real-world applicability of BioMARL, we analyzed cancer patient data from Vanderbilt University Medical Center. Compared to the widely used *t*-test, a fundamental statistical method for biomarker identification in cancer studies [42], BioMARL-selected genes demonstrated stronger biological relevance, with significant associations with key cancer pathways and greater therapeutic potential, as evidenced by their prevalence in high-impact cancer journals. These results demonstrate the robustness of BioMARL and highlight its potential to advance biomarker-driven precision oncology by improving drug development and clinical trial design.

5 Related Work

Gene Selection in Genomic Studies. Feature engineering algorithms for gene expression data fall into three categories: filter, wrapper, and embedded methods. Filter methods [17, 18, 34] evaluate and filter features based on their potential contribution to model performance, typically using relevance scores and thresholds for gene selection. For example, the Grouping Genetic Algorithm [18] addresses feature grouping in RNA-Seq data. Wrapper methods [2, 29] assess feature significance using classification algorithms like k-Nearest Neighbors, Random Forests, and Support Vector Machines. Embedded methods [26, 54] combine classifier configuration with feature subset exploration. However, these traditional approaches often neglect crucial gene interactions and biological context, while struggling to handle the dynamic nature of gene expression data, potentially missing key biomarkers.

Multi-agent Reinforcement Learning (MARL). MARL addresses multi-agent environments [40] where agents interact to achieve shared or individual goals. The combinatorial nature and environmental complexity [31] make training challenging, with many MARL applications addressing NP-Hard problems like manufacturing scheduling [14, 16], vehicle routing [39, 51], and multi-agent games [6, 36]. While deep RL has advanced to handle large-scale systems [47], and both multi-agent [4, 19, 30, 46, 49, 50] and single-agent [44, 55] approaches have been developed for feature selection, these methods lack comprehensive domain knowledge integration and are not optimized for HDLSS data scenarios, where direct application without proper dimensionality reduction can lead to computational inefficiency and poor convergence.

6 Concluding Remarks

We introduce *BioMARL*, a novel framework that transforms gene selection into a pathway-aware collaborative decision process for genomic analysis. By unifying statistical feature selection with biological pathway knowledge through MARL, our approach bridges the divide between purely statistical methods and biological context in gene selection. The framework's two-stage architecture, combining pathway-guided pre-filtering with collaborative agent modeling, ensures both statistical rigor and biological relevance in selected gene signatures. Through extensive evaluation on multiple TCGA datasets, BioMARL demonstrates superior predictive performance while maintaining interpretability through pathway-level insights. Gene Ontology enrichment analysis confirms that selected genes form biologically coherent signatures, validating our

framework's ability to capture meaningful pathway interactions. These genes exhibit distinct expression patterns that differentiate breast cancer patients, highlighting the framework's potential for biological discovery and its applicability in precision oncology.

7 Acknowledgement

This research was partially supported by the National Science Foundation (NSF) via the grant numbers: 2152030, 2246796, 2426340, 2416727, 2421864, 2421865, 2421803, and National Academy of Engineering Grainger Foundation Frontiers of Engineering Grants.

References

- [1] Khandakar Tanvir Ahmed, Jiao Sun, Sze Cheng, Jeongsik Yong, and Wei Zhang. 2022. Multi-omics data integration by generative adversarial network. *Bioinformatics* 38, 1 (2022), 179–186.
- [2] Fuad Al Abir, SM Shovan, Md Al Mehedi Hasan, Abu Sayeed, and Jungpil Shin. 2022. Biomarker identification by reversing the learning mechanism of an autoencoder and recursive feature elimination. *Molecular Omics* 18, 7 (2022), 652–661.
- [3] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. 2000. Gene ontology: tool for the unification of biology. *Nature genetics* 25, 1 (2000), 25–29.
- [4] Ehtesamul Azim, Dongjie Wang, Kunpeng Liu, Wei Zhang, and Yanjie Fu. 2024. Feature Interaction Aware Automated Data Representation Transformation. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*. SIAM, 878–886.
- [5] Oluleye H Babatunde, Leisa Armstrong, Jinsong Leng, and Dean Diepeveen. 2014. A genetic algorithm-based feature selection. (2014).
- [6] Nolan Bard, Jakob N Foerster, Sarah Chandar, Neil Burch, Marc Lanctot, H Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, et al. 2020. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence* 280 (2020), 103216.
- [7] Sudipto Baul, Khandakar Tanvir Ahmed, Joseph Filipek, and Wei Zhang. 2022. omicsGAT: Graph attention network for cancer subtype analyses. *International Journal of Molecular Sciences* 23, 18 (2022), 10220.
- [8] William W Bivin, Oleksandr Yergiyev, Mark L Bunker, Jan F Silverman, and Uma Krishnamurti. 2017. GRB7 expression and correlation with HER2 amplification in invasive breast carcinoma. *Applied Immunohistochemistry & Molecular Morphology* 25, 8 (2017), 553–558.
- [9] Brigham & Women's Hospital and others. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 7418 (2012), 61–70.
- [10] Cancer Genome Atlas Research Network. 2013. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 10 (2013), 1113–1120.
- [11] Cancer Genome Atlas Research Network. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 7511 (2014), 543.
- [12] Cancer Genome Atlas Research Network and others. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 7353 (2011), 609.
- [13] Han-Yu Chuang, Eunjung Lee, Yu-Tsung Liu, Doheon Lee, and Trey Ideker. 2007. Network-based classification of breast cancer metastasis. *Molecular systems biology* 3, 1 (2007), 140.
- [14] Marc-André Dittrich and Silas Fohlmeister. 2020. Cooperative multi-agent system for production control using reinforcement learning. *CIRP Annals* 69, 1 (2020), 389–392.
- [15] Aileen I Fernandez, Matthew Liu, Andrew Bellizzi, Jane Brock, Oluwole Fadare, Krisztina Hanley, Malini Harigopal, Julie M Jorns, M Gabriela Kuba, Amy Ly, et al. 2022. Examination of low ERBB2 protein expression in breast cancer tissue. *JAMA oncology* 8, 4 (2022), 607–610.
- [16] Thomas Gabel and Martin Riedmiller. 2007. On a successful application of multi-agent reinforcement learning to operations research benchmarks. In *2007 IEEE international symposium on approximate dynamic programming and reinforcement learning*. IEEE, 68–75.
- [17] Consolata Gakii, Paul O Mireji, and Richard Rimuru. 2022. Graph based feature selection for reduction of dimensionality in next-generation rna sequencing datasets. *Algorithms* 15, 1 (2022), 21.
- [18] Pilar García-Díaz, Isabel Sánchez-Berriel, Juan A Martínez-Rojas, and Ana M Diez-Pascual. 2020. Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. *Genomics* 112, 2 (2020), 1916–1925.
- [19] Nanxu Gong, Wangyang Ying, Dongjie Wang, and Yanjie Fu. 2025. Neuro-symbolic embedding for short and effective feature selection via autoregressive generation. *ACM Transactions on Intelligent Systems and Technology* 16, 2 (2025), 1–21.

[20] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46 (2002), 389–422.

[21] Ryan J Hartmaier, Lee A Albacker, Juliani Chmielecki, Mark Bailey, Jie He, Michael E Goldberg, Shakti Ramkissoon, James Suh, Julia A Elvin, Samuel Chiacchia, et al. 2017. High-throughput genomic profiling of adult solid tumors reveals novel insights into cancer pathogenesis. *Cancer research* 77, 9 (2017), 2464–2475.

[22] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. 2022. Ensemble of feature selection algorithms: a multi-criteria decision-making approach. *International Journal of Machine Learning and Cybernetics* 13, 1 (2022), 49–69.

[23] Nichola Johnson, Paolo De Ieso, Gabriele Migliorini, Nick Orr, Peter Broderick, Daniel Catovsky, Athena Matakidou, Timothy Eisen, Christy Goldsmith, Frank Dudbridge, et al. 2016. Cytochrome P450 allele CYP3A7* 1C associates with adverse outcomes in chronic lymphocytic leukemia, breast, and lung cancer. *Cancer research* 76, 6 (2016), 1485–1493.

[24] Minoru Kanehisa and Susumu Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28, 1 (2000), 27–30.

[25] Tzu-Jen Kao, Chung-Chu Wu, Nam Nhu Phan, Yen-Hsi Liu, Hoang Dang Khoa Ta, Gangga Anuraga, Yung-Fu Wu, Kuen-Haur Lee, Jian-Ying Chuang, and Chih-Yang Wang. 2021. Prognoses and genomic analyses of proteasome 26S subunit, ATPase (PSMC) family genes in clinical breast cancer. *Aging (Albany NY)* 13, 14 (2021), 17970.

[26] Yunchuan Kong and Tianwei Yu. 2018. A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data. *Bioinformatics* 34, 21 (2018), 3727–3737.

[27] Patrick Lebok, Melike ÖzTÜRK, Uwe Heilenkötter, Fritz Jaenische, Volkmar Müller, Peter Paluchowski, Stefan Geist, Christian Wilke, Eicke Burandt, Annette Lebeau, et al. 2016. High levels of class III β -tubulin expression are associated with aggressive tumor features in breast cancer. *Oncology Letters* 11, 3 (2016), 1987–1994.

[28] Ismael Lemhadri, Feng Ruan, Louis Abraham, and Robert Tibshirani. 2021. Lassonet: A neural network with feature sparsity. *Journal of Machine Learning Research* 22, 127 (2021), 1–29.

[29] Yuanyuan Li, Kai Kang, Juno M Krahm, Nicole Croutwater, Kevin Lee, David M Umbach, and Leping Li. 2017. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC genomics* 18 (2017), 1–13.

[30] Kunpeng Liu, Yanjie Fu, Pengfei Wang, Le Wu, Rui Bo, and Xiaolin Li. 2019. Automating feature subspace exploration via multi-agent reinforcement learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 207–215.

[31] Kunpeng Liu, Haibo Huang, Wei Zhang, Ahmad Hariri, Yanjie Fu, and Kien Hua. 2021. Multi-armed bandit based feature selection. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*. SIAM, 316–323.

[32] Catalina Lodillinsky, Laetitia Fuhrmann, Marie Ironnelle, Olena Pylypenko, Xiao-Yan Li, Hélène Bonsang-Kitzis, Fabien Reynal, Sophie Vacher, Claire Calmel, Olivier De Wever, et al. 2021. Metastasis-suppressor NME1 controls the invasive switch of breast cancer by regulating MT1-MMP surface clearance. *Oncogene* 40, 23 (2021), 4019–4032.

[33] Andrew Mulcahy, Stephanie Renname, Daniel Schwam, Reid Dickerson, Lawrence Baker, and Kanaka Shetty. 2025. Use of Clinical Trial Characteristics to Estimate Costs of New Drug Development. *JAMA Network Open* 8, 1 (2025), e2453275–e2453275.

[34] Sungsoo Park, Bonggun Shin, Won Sang Shim, Yoonjung Choi, Kilsoo Kang, and Keunsoo Kang. 2019. Wx: a neural network-based feature selection algorithm for transcriptomic data. *Scientific reports* 9, 1 (2019), 10500.

[35] Hanchuan Peng, Fuhui Long, and Chris Ding. 2005. 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 (2005), 1226–1238.

[36] Peng Peng, Ying Wen, Yaodong Yang, Quan Yuan, Zhenkun Tang, Haitao Long, and Jun Wang. 2017. Multiagent bidirectionally-coordinated nets: Emergence of human-level coordination in learning to play starcraft combat games. *arXiv preprint arXiv:1703.10069* (2017).

[37] Daniel R Rhodes, Bushra Ateeq, Qi Cao, Scott A Tomlins, Rohit Mehra, Bharathi Laxman, Shanker Kalyana-Sundaram, Robert J Lonigro, Beth E Helgeson, Mahaveer S Bhojani, et al. 2009. AGTR1 overexpression defines a subset of breast cancer and confers sensitivity to losartan, an AGTR1 antagonist. *Proceedings of the National Academy of Sciences* 106, 25 (2009), 10284–10289.

[38] Borja Seijo-Pardo, Iago Porto-Díaz, Verónica Bolón-Canedo, and Amparo Alonso-Betanzos. 2017. Ensemble feature selection: Homogeneous and heterogeneous approaches. *Knowledge-Based Systems* 118 (2017), 124–139.

[39] Maria Amélia Lopes Silva, Sérgio Ricardo de Souza, Marcone Jamilson Freitas Souza, and Ana Lúcia C Bazzan. 2019. A reinforcement learning-based multi-agent framework applied for solving routing and scheduling problems. *Expert Systems with Applications* 131 (2019), 148–171.

[40] Ardi Tampuu, Tammet Matiisen, Dorian Kodelja, Ilya Kuzovkin, Kristjan Korus, Juhan Aru, Jaan Aru, and Raul Vicente. 2017. Multiagent cooperation and competition with deep reinforcement learning. *PLoS one* 12, 4 (2017), e0172395.

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

[42] Laura J Van't Veer, Hongyu Dai, Marc J Van De Vijver, Yudong D He, Augustinus AM Hart, Mao Mao, Hans L Peterse, Karin Van Der Kooy, Matthew J Marton, Anke T Witteveen, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *nature* 415, 6871 (2002), 530–536.

[43] Meng Wang, Guangxin Zhang, Yue Zhang, Xuelian Cui, Shuaibin Wang, Song Gao, Yicun Wang, Ying Liu, Jeeyoo H Bae, Wei-Hsiung Yang, et al. 2020. Fibrinogen alpha chain knockout promotes tumor growth and metastasis through integrin-AKT signaling pathway in lung cancer. *Molecular Cancer Research* 18, 7 (2020), 943–954.

[44] Xinyuan Wang, Dongjie Wang, Wangyang Ying, Rui Xie, Haifeng Chen, and Yanjie Fu. 2024. Knockoff-Guided Feature Selection via A Single Pre-trained Reinforced Agent. *arXiv preprint arXiv:2403.04015* (2024).

[45] Yixin Wang, Jan GM Klijn, Yi Zhang, Anieta M Sieuwerts, Maxime P Look, Fei Yang, Dmitri Talantov, Mieke Timmermans, Marion E Meijs-van Gelder, Jack Yu, et al. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* 365, 9460 (2005), 671–679.

[46] Meng Xiao, Dongjie Wang, Min Wu, Pengfei Wang, Yuanchun Zhou, and Yanjie Fu. 2023. Beyond discrete selection: Continuous embedding space optimization for generative feature selection. *arXiv preprint arXiv:2302.13221* (2023).

[47] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. 2018. Mean field multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 5571–5580.

[48] Yiming Yang, Jan O Pedersen, et al. 1997. A comparative study on feature selection in text categorization. In *icml*, Vol. 97. Citeseer, 35.

[49] Wangyang Ying, Dongjie Wang, Haifeng Chen, and Yanjie Fu. 2024. Feature selection as deep sequential generative learning. *ACM Transactions on Knowledge Discovery from Data* 18, 9 (2024), 1–21.

[50] Wangyang Ying, Dongjie Wang, Xuanming Hu, Ji Qiu, Jin Park, and Yanjie Fu. 2024. Revolutionizing Biomarker Discovery: Leveraging Generative AI for Bio-Knowledge-Embedded Continuous Space Exploration. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 5046–5053.

[51] Ke Zhang, Fang He, Zhengchao Zhang, Xi Lin, and Meng Li. 2020. Multi-vehicle routing problems with soft time windows: A multi-agent reinforcement learning approach. *Transportation Research Part C: Emerging Technologies* 121 (2020), 102861.

[52] Wei Zhang, Jeremy Chien, Jeongsik Yong, and Rui Kuang. 2017. Network-based machine learning and graph theory algorithms for precision oncology. *NPJ precision oncology* 1, 1 (2017), 25.

[53] Wei Zhang, Nicholas Johnson, Baolin Wu, and Rui Kuang. 2012. Signed network propagation for detecting differential gene expressions and DNA copy number variations. In *Proceedings of the ACM conference on bioinformatics, computational biology and biomedicine*. 337–344.

[54] Zishuang Zhang and Zhi-Ping Liu. 2021. Robust biomarker discovery for hepatocellular carcinoma from high-throughput data by multiple feature selection methods. *BMC medical genomics* 14 (2021), 1–12.

[55] Xiaosa Zhao, Kunpeng Liu, Wei Fan, Lu Jiang, Xiaowei Zhao, Minghao Yin, and Yanjie Fu. 2020. Simplifying reinforced feature selection via restructured choice strategy of single agent. In *2020 IEEE International conference on data mining (ICDM)*. IEEE, 871–880.

8 Appendix

8.1 Implementation Details

Hyperparameters and Reproducibility

Pathway-Guided Gene Pre-filtering: For each pathway, we evaluate performance using a Random Forest classifier with 100 trees and default scikit-learn parameters. The pathway performance bonus scaling factor β is set to 0.2.

Multi-Agent Gene Selection Framework: Each agent's DQN consists of a 4-layer network (256-128-64-2 neurons) with ReLU activation and layer normalization. The learning rate is 0.0003, the discount factor γ is 0.85, the initial ϵ is 0.95 with a decay rate of 0.99, and the minimum ϵ is 0.1. The experience replay buffer capacity is 1700 with 3000 exploration steps for each run, with a batch size of 64. Target network updates occur every 50 steps. The Graph Neural Network uses two graph convolutional layers with a hidden dimension of

Algorithm 1 Pathway-Aware Gene Selection Framework

Input: Dataset D with gene set G , KEGG pathway database P , Maximum features k
Output: Selected gene set G_{opt} with $|G_{opt}| = k$

- 1: **Phase 1: Pathway-Guided Pre-filtering (Section 3.2)**
- 2: **for all** method $f_i \in \{\text{chi-squared, random forest, SVM}\}$ **do**
- 3: Compute scores $S_i \leftarrow f_i(D)$ for all genes
- 4: $w_i \leftarrow \text{performance}_i / \sum_j \text{performance}_j$
- 5: **end for**
- 6: **for all** pathway $p_i \in P$ **do**
- 7: $S_p(p_i) \leftarrow \text{classifier_performance}(D[G_i])$
- 8: **end for**
- 9: **for all** gene $g \in G$ **do**
- 10: $m_g \leftarrow \sum_i w_i \cdot S_{i,g}$
- 11: $\hat{s}_g \leftarrow m_g \cdot (1 + \beta \cdot \log(1 + S_p(g)))$
- 12: **end for**
- 13: $G_{pre} \leftarrow \{g \in G : \hat{s}_g > \mu + 2\sigma\}$
- 14: **Phase 2: Multi-Agent Gene Selection (Section 3.3)**
- 15: Initialize DQN agents for each gene in G_{pre}
- 16: Initialize shared memory M with synergy matrix M_{ij} and success record C
- 17: Initialize centralized critic V with gating mechanism
- 18: Initialize meta-learner ensemble f_{meta}
- 19: **for** episode = 1 to E **do**
- 20: $s_t \leftarrow \text{GNN}(X_t, E_t)$ with pathway-augmented edges
- 21: **for** $t = 1$ to T **do**
- 22: **for all** agent i **do**
- 23: Get top partners P_i from M_{ij}
- 24: $\text{bias}_i \leftarrow \beta \sum_{j \in P_i} M_{ij}$
- 25: $P(a_i = 1) \leftarrow \sigma(Q_i(s_t, 1) + \text{bias}_i)$
- 26: Select a_i using ϵ -greedy policy
- 27: **end for**
- 28: // Performance estimation
- 29: $D_t \leftarrow \text{construct perturbation matrix}$
- 30: $\Delta P_t, u_t \leftarrow f_{meta}(D_t)$ ▷ Changes and uncertainties
- 31: $c_t \leftarrow 1/(1 + u_t)$ ▷ Confidence factors
- 32: $r_{base} \leftarrow \Delta P_t \odot c_t - \log(1 + u_t)$
- 33: // Pathway metrics
- 34: **for all** gene i **do**
- 35: **if** $a_i = 0$ **then**
- 36: $\Delta\phi_i \leftarrow \Phi(S \cup \{i\}) - \Phi(S)$
- 37: $\Delta\psi_i \leftarrow \psi_p(S \cup \{i\}) - \psi_p(S)$
- 38: **else**
- 39: $\Delta\phi_i \leftarrow \Phi(S) - \Phi(S \setminus \{i\})$
- 40: $\Delta\psi_i \leftarrow \psi_p(S) - \psi_p(S \setminus \{i\})$
- 41: **end if**
- 42: $r_i \leftarrow \alpha r_{base} + \beta \Delta\phi_i + \gamma \Delta\psi_i$
- 43: **end for**
- 44: // Critic and memory updates
- 45: $v_t \leftarrow V(s_t)$ ▷ Global value estimate
- 46: Update M_{ij} for selected gene pairs
- 47: Decay memory weights: $M_{ij} \leftarrow \lambda M_{ij}$
- 48: **for all** agent i **do**
- 49: $Q_{target} \leftarrow \lambda_a(r_i + \gamma \max_{a'} Q_i(s_{t+1}, a')) + \lambda_b v_t$
- 50: Update DQN using $(s_t, a_i, r_i, s_{t+1}, Q_{target})$
- 51: **end for**
- 52: Update critic by minimizing $\|v_t - I_t\|^2$
- 53: $s_t \leftarrow s_{t+1}$
- 54: **end for**
- 55: **end for**
- 56: Compute final weighted scores using Q -value differences
- 57: Rank genes based on weighted scores to get G_{ranked}
- 58: $G_{opt} \leftarrow \text{top } k \text{ genes from } G_{ranked}$
- 59: **return** G_{opt}

64, and pathway embeddings of dimension 64. Edge creation uses a correlation weight α of 0.7. λ_a and λ_b of the critic was set to 0.7 and 0.3 respectively. The meta-learner combines Random Forest, XGBoost, and LightGBM regressors (each with 100 estimators), with a linear meta-model and a neural network (256-128-64-32-1 neurons). The final reward weighting coefficients (ω, ξ, ζ) were set to 0.5, 0.25, and 0.25 respectively. Online updates of the meta learner occur every 50 steps. The shared memory mechanism uses a decay factor of 0.99 and an exploration bonus of 0.005. The synergy bias weight β is initialized at 0.08 and increases linearly up to 0.3. All experiments were repeated 10 times with different random seeds.

Environmental Settings. All experiments are conducted on the Ubuntu 22.04.5 LTS operating system, AMD Ryzen Threadripper 2950X 16-Core Processor, and 3 NVIDIA RTX A4500 GPUs (20GB VRAM each) with 128GB of system RAM. The framework uses Python 3.10.12 and PyTorch 2.2.1 with CUDA 11.8 support.

8.2 Pseudocode of the algorithm

The entire procedure algorithm is described as Algorithm 1.