**Proceedings Paper:**
Tabakhi, S. and Lu, H. (2022) Multi-agent feature selection for integrative multi-omics analysis. In: Proceedings of 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 11-15 Jul 2022, Glasgow, Scotland. Institute of Electrical and Electronics Engineers (IEEE), pp. 1638-1642. ISBN: 9781728127835. ISSN: 2375-7477. EISSN: 2694-0604.

https://doi.org/10.1109/EMBC48229.2022.9871758

# Multi-agent Feature Selection for Integrative Multi-omics Analysis

Sina Tabakhi, *Student Member*, *IEEE*, and Haiping Lu, *Senior Member*, *IEEE*

***Abstract*— Multi-omics data integration is key for cancer prediction as it captures different aspects of molecular mechanisms. Nevertheless, the high-dimensionality of multi-omics data with a relatively small number of patients presents a challenge for the cancer prediction tasks. While feature selection techniques have been widely used to tackle the curse of dimensionality of multi-omics data, most existing methods have been applied to each type of omics data separately. In this paper, we propose a multi-agent architecture for feature selection, called MAgentOmics, to consider all omics data together. MAgentOmics extends the ant colony optimization algorithm to multi-omics data, which iteratively builds candidate solutions and evaluates them. Moreover, a new fitness function is introduced to assess the candidate feature subsets without using prediction target such as survival time of patients. Therefore, it can be considered as an unsupervised method. We evaluate the performance of MAgentOmics on the TCGA ovarian cancer multi-omics data from 176 patients using a 5-fold cross-validation. The results demonstrate that the integration power of MAgentOmics is relatively better than the state-of-the-art supervised multi-view method. The code is publicly available at `https://github.com/SinaTabakhi/MAgentOmics`.**

***Clinical relevance*— Discovering knowledge in existing multi-omics datasets through better feature selection enhances the clinical understanding of cancers and speeds-up decision-making in the clinic.**

## I. INTRODUCTION

The advent of high-throughput technologies has recently generated massive amounts of biological omics data, provided the ability of comprehensive assessment of a set of molecules, and created new trends for biologists and data scientists to diagnose, treat and even cure cancers [1], [2]. Each type of omics data provides unique insights into biological processes. Multi-omics integration aims to simultaneously and comprehensively measure these biological molecules to obtain a deep understating of complex molecular mechanisms that lead to diseases [3], [4]. Public multi-omics datasets such as The Cancer Genome Atlas (TCGA) [5] and International Cancer Genomics Consortium (ICGC) [6] programs have collected comprehensive profiles of several cancer types for multiple molecular layers, encouraging researchers to study cancer-causing features. Therefore, the integration task is a challenging issue for researchers working with high multi-dimensionality datasets with a relatively small number of samples. This phenomenon is called "the curse of dimensionality," presenting many challenges to machine learning tasks [7]–[9].

Sina Tabakhi and Haiping Lu are with the Department of Computer Science, the University of Sheffield, United Kingdom {stabakhi1,h.lu@sheffield.ac.uk}

A way to categorize existing integration methods for different machine learning tasks is needed to select best practices. Multi-omics integration methods have been widely classified into early, late, and intermediate integration approaches [10]. The early integration approach considers the concatenation of each omics dataset into a single joint matrix, and machine learning models are applied on the produced single matrix. Support vector machines [11] and deep neural networks [12] have been used to analyze combined multi-omics data. Though the early integration is simple and easy to implement, it increases the number of features and ignores the size difference among omics data which makes the learning process difficult [10].

In comparison, the late integration approach employs machine learning models on each omics data separately and then aggregates models results. MOGONET [13] is a complex example of late integration for the biomedical classification task. This approach can use available methods that have been proposed explicitly for each omics type. Ignoring interactions among omics data to utilize the complementarity information between them is the main disadvantage of the late integration approach [7], [10].

On the other hand, the intermediate integration approach contains methods that construct a single joint model of the multi-omics data. Huang et al. [14] proposed deep learning-based networks to aggregate different types of omics data for survival analysis of breast cancer. Generally, this approach outperforms other approaches because it assumes that the different omics data share a common space which shows the underlying biological mechanisms [15].

Since multi-omics data possess a high dimensionality, many studies have utilized feature selection to simplify the integration process [16]. However, feature selection methods have been independently applied to each omics data as a preprocessing step which neglects all omics data together. Therefore, the genuine joint feature selection for multi-omics data remains very rare. EL-Manzalawy et al. [17] proposed a multi-view feature selection based on min-redundancy and max-relevance criteria for multi-omics data, called mRMR-mv. They used an incremental feature selection process to select a feature subset from all views with maximal relevance to the target class and minimal redundancy between pairs of selected features. However, their method identifies the subset of features in a single-iteration process that is started from a specific point; as a result, it can be easily trapped into the local optimum.

Multi-agent systems (MAS) consist of multiple simple agents collaborating in a shared environment to achieve an objective. Among many MAS algorithms, ant colony

optimization (ACO) has extensively been used for feature selection on single view data due to their acceptable performance [18]–[20]. The main superiority of the ACO compared to other MAS algorithms can be noted as the availability of a distributed long-term memory to share knowledge between ants, simulation of reinforcement learning concepts for finding better solutions, and parallel nature of the algorithm to reduce the execution time [9], [21].

This study aims to design a multi-agent architecture for multi-view (i.e. multi-omics) feature selection to consider different omics data together, named MAgentOmics. To the best of our knowledge, it is the first attempt to apply a MAS for multi-view feature selection to handle the high-dimensionality nature of multi-omics data. In the proposed method, the labels of samples are not required in the feature selection process, and it can be considered as an unsupervised method. MAgentOmics consists of ants that interact with each other through sharing knowledge. Iteratively, each ant generates a subset of features using a combination of relevance and redundancy analyses. Then, a new fitness function is proposed to evaluate the constructed solutions. To incorporate different views in the feature selection process, a probability distribution is defined to show the relative importance of the view, and it gets updated through a new probability updating rule. Finally, the global-best solution is chosen as the final feature subset.

## II. MATERIALS AND METHODS

### A. Multi-omics Data

The performance of the MAgentOmics method has been evaluated on a publicly available dataset from the TCGA program [22]. We selected ovarian serous cystadenocarcinoma data from the TCGA and analyzed three omics, including gene-level copy number variation, DNA methylation, and gene expression RNAseq. The ovarian data were downloaded from the UCSC Xena Platform[1]. A brief description of the dataset is shown in Table I.

For the classification task, the data were divided into two groups based on the clinical information (i.e. recorded time of death and vital status) in the TCGA ovarian cancer, namely, long-term and short-term groups. The long-term group includes those samples with survival time $\geq$ 3 years. On the other hand, the short-term group includes samples with survival time $<$ 3 years and vital status of 'DECEASED' [17]. Since data from all omics were not measured for each patient, only those patients with all omics measured and available clinical information are included in this study. Therefore, the final multi-omics data used further are 176 samples, 85 of which belong to the long-term group.

### B. Data Pre-processing

Raw data of ovarian cancer were then pre-processed by the following steps to ensure the robustness for computation in the classification task [22], [23]. Features with missing values in each omics data were removed. For each feature in the data, then, values were normalized to the range of [0, 1] by applying the min-max normalization approach [24]. Finally, retained features with variance values lower than 0.05 were filtered out.

[1]https://xenabrowser.net/datapages/.

TABLE I
CHARACTERISTICS OF THE TCGA OVARIAN CANCER MULTI-OMICS DATA USED IN THE EXPERIMENTS.

| Omics Type | Version | #Features | #Samples |
|---|---|---|---|
| DNA methylation | 2017-09-08 | 27,578 | 616 |
| Gene-level copy number variation | 2017-09-08 | 24,776 | 579 |
| Gene expression RNA-seq | 2017-10-13 | 20,530 | 308 |

### C. Multi-agent Feature Selection Architecture

MAgentOmics is a multi-agent architecture for feature selection designed for multi-omics data. The overall framework of the proposed method to find the near-optimal feature subset is comprised of the ACO algorithm. Therefore, the search space should be modeled as a suitable graph for ACO before starting the feature selection procedure, which is illustrated in Fig. 1.

More formally, each omics data can be represented as a complete weighted graph, $G =< F^v, E^v, p_v >$, where $F^v$ denotes the nodes in the graph corresponded to a set of available features in $v$-th view, and $E^v = \{(f_i^v, f_j^v), f_i^v, f_j^v \in F^v\}$ indicates the set of edges of the graph. Each edge is assigned a value corr$(f_i^v, f_j^v)$, which is the correlation between features $f_i^v$ and $f_j^v$. To compute the correlation between each pair of features, we use the absolute value of the Pearson's correlation coefficient between them [24]. Additionally, $p_v$ denotes the global relative importance of $v$-th view in the multi-omics data that estimates the probability of selecting a feature from $v$-th view.

In the ACO algorithm used in the feature selection problem, ants build feature subsets by moving from feature to feature on the graph. The ants' solution construction is guided by two components: "pheromone" and "heuristic information". A pheromone strength $\tau_i^v$ is associated with the features and will be changed during the search process to show the information obtained by ants. The heuristic information is used as the prior knowledge of the problem to guide the search strategy of the ACO algorithm. In this paper, two static heuristic information are defined, which are calculated at the initialization time and do not change during the algorithm's run. The first heuristic function is computed as the inverse of the correlation between features. Moreover, each feature's relevance is considered as the second heuristic function and measured by the term variance criterion [24].

Algorithm 1 describes the framework of the proposed multi-agent feature selection method. This framework is comprised of three main sections, including initialization (lines 1-4), feature selection procedure (lines 5-18), and final dataset construction (lines 19-20).

Initially, the correlation values between pairs of features from each view are computed and assigned to the graph edges. Then, the relevance of each feature is separately assessed by utilizing the term variance. Thereafter, the initial value of pheromone on each feature is set to a small constant

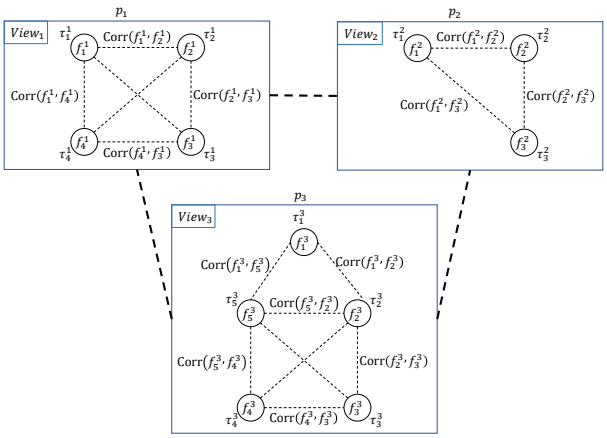



Fig. 1. The graph representation of the search space for the multi-view feature selection problem.

$c$. Finally, the initial probability values are equally distributed among views ($p_k = 1/v, \forall k = 1, 2, \ldots, v$).

The feature selection procedure in MAgentOmics involves iterative improvement in which any iteration of the algorithm works in five steps.

**Step 1**: For each view $v$, $N_A$ ants are put on a randomly chosen graph node from view $v$ as their initial states.

**Step 2**: At the construction step, each ant extends its current partial solution by sequentially employing "state transition rules" biased by the amount of pheromone and heuristic information. Ants, generally, prefer features that are marked by the more substantial pheromone trail and higher relevancy, as well as have a lower correlation with the previously selected feature. These rules allow ants to greedily search for the following features at their current view or probabilistically explore other views for prospective features. The subset generation will be stopped when the ants select a pre-defined number of features.

**Step 3**: The quality of the generated candidate subsets is measured using a new fitness function. Then, the best subset found at the current iteration is kept as the current-best solution.

**Step 4**: Once the solutions have been built and evaluated, pheromone trails can be updated through a process called the "pheromone updating rule." This process acts by the fact that the pheromone evaporates after some time, and less promising features progressively lose pheromone. On the other hand, the features which are repeatedly selected by the ants are more likely to be chosen in the subsequent iterations and get the greater level of pheromone. Furthermore, the ant with the current-best solution can deposit an extra amount of pheromone on its selected features.

**Step 5**: The probability distribution is updated by the concept of reinforcement learning; an agent is encouraged to explore the state that has obtained a high positive reinforcement in the long run [25]. To simulate this concept, the number of features selected by ants in a specific view is relatively considered as the importance of that view.

These five steps are repeated until the maximum number of iterations is met. The global-best solution is selected among

**Algorithm 1** MAgentOmics – Multi-agent Feature Selection for Multi-view Data Integration

**Input**
$\mathbb{D} =< (X^1, X^2, ..., X^v), y >$: multi-view dataset.
$N_I$: maximum number of iterations.
$N_A$: number of agents placed in each view.
**Output**
$\mathbb{D}' =< X', y >$: final single dataset $X', d' \times n$.

Calculate $\mathrm{corr}(f_i^k, f_j^k), \forall k = 1, 2, ..., v$.
Calculate $\mathrm{rel}(f_i^k), \forall k = 1, 2, ..., v$.
$\tau_i^k(0) \leftarrow c, \forall k = 1, 2, ..., v$. ▷ Initialize pheromone
$p_k \leftarrow \frac{1}{v}, \forall k = 1, 2, ..., v$. ▷ Initialize probability
**for** $t = 1$ to $N_I$ **do**
  **for** $k = 1$ to $v$ **do**
    Put $N_A$ agents on a randomly chosen node.
  **end for**
  **for** $k = 1$ to $v$ **do**
    **for** $a = 1$ to $N_A$ **do**
      Form new feature subset by (1), (2) or (3).
      Evaluate the generated subset by (6).
    **end for**
  **end for**
  Select the current-best solution at $t$-th iteration.
  Update the pheromone values by applying (4).
  Update probability distribution by applying (7).
**end for**
Choose the global-best solution found.
Construct $\mathbb{D}'$ based on the global-best solution.

the best solutions in all iterations. Each omics data is then reduced based on the features represented in the global-best solution. The final dataset is constructed by combining multiple reduced omics datasets.

*State Transition Rule*. The selection of the next feature to add to the solution is made greedily or probabilistically at each construction step. In the greedy rule, the maximum value with which an ant, currently at feature $i$ from view $v$, decides to select feature $j$ from the same view at the $t$-th iteration of the algorithm is:

$$f_j^v = \underset{j \in \Omega_{i^v}^v}{\arg\min} \left[ \tau_j^v(t) \left[ \eta_1(f_j^v) \right]^\alpha \left[ \eta_2(f_i^v, f_j^v) \right]^\beta \right] \text{ if } q \leq q_0, \quad (1)$$

where $\Omega_{i^v}^v$ is unvisited features in view $v$ for the ant, $\tau_j^v(t)$ is the pheromone level on feature $j$ from view $v$ at time $t$, $\eta_1(f_j^v)$ is the first heuristic information which shows the relevance of feature $j$ from view $v$, $\eta_2(f_i^v, f_j^v)$ is the second heuristic information which indicates the inverse of the correlation values between features $f_i^v$ and $f_j^v$, the role of parameters $\alpha$ and $\beta$ is to determine the relative importance of pheromone and two heuristic information values, $q$ is a uniform random number distributed in $[0, 1]$ and $q_0$ is a parameter ($0 \leq q_0 \leq 1$).

In the probabilistic way, first of all, an ant placed in view $v$ chooses a random view $v'$ from the set of views $V$ ($v'$ can be equal to $v$) by considering the probability distribution of views $\mathbb{P} = \{p_1, p_2, \ldots, p_v\}$ as follows:

$$v' = \text{choice}(V, \mathbb{P}) \quad \text{if} \quad q > q_0. \quad (2)$$

Once the next view has been chosen, the next feature $j$ from view $v'$ is selected based on the probability $P(f_j^{v'} | f_i^v)$,

which is calculated according to the following equation:

$$P(f_j^{v'}|f_i^v) = \frac{\tau_j^{v'}(t)\left[\eta_1\left(f_j^{v'}\right)\right]^{\alpha}\left[\eta_2\left(f_i^v, f_j^{v'}\right)\right]^{\beta}}{\sum\limits_{u\in\Omega_{iv}^{v'}} \tau_u^{v'}(t)\left[\eta_1\left(f_u^{v'}\right)\right]^{\alpha}\left[\eta_2\left(f_i^v, f_u^{v'}\right)\right]^{\beta}}. \quad (3)$$

*Pheromone Updating Rule*. After all ants have formed candidate subsets, the intensity of pheromone level on the nodes is updated, which is commonly implemented as:

$$\tau_i^v(t+1) = (1-\rho)\tau_i^v(t) + \rho\left[\frac{\text{count}(\{f_i^v\})}{\text{count}(S)} + \left[\Delta\tau_i^v(t)\right]^{\text{best}}\right], \quad (4)$$

where

$$\left[\Delta\tau_i^v(t)\right]^{\text{best}} = \begin{cases} [\text{fitness}(S_a)]^{\text{best}} & \text{if } f_i^v \in S_a, \\ 0 & \text{otherwise}, \end{cases} \quad (5)$$

$\rho$ is a parameter that is called evaporation rate ($0 < \rho < 1$), $\tau_i^v(t)$ and $\tau_i^v(t+1)$ indicate the pheromone level on feature $i$ from view $v$ at time $t$ and $t+1$, respectively, count(.) is a function that counts the number of times that a subset of features have been selected during the current iteration, and $S$ is the subset of all selected features during the current iteration. In Eq. (5), fitness$[(S_a)]^{\text{best}}$ denotes a fitness function which evaluates the quality of the constructed subset $S_a$ by ant $a$, and 'best' notation shows the current-best solution.

*Fitness Function*. The goal of the fitness function is to figure out the quality of a feature subset. In this study, the fitness function is designed to incorporate two important metrics in the feature selection research area, maximal relevance and minimal redundancy, which is defined as:

$$\text{fitness}(S_a) = \frac{\sum_{f_i^v \in S_a} \text{rel}(f_i^v)/|S_a|}{\sum_{f_i^v, f_j^{v'} \in S_a} \text{corr}(f_i^v, f_j^{v'}) / \sum_{i=1}^{|S_a|-1} i}, \quad (6)$$

where $\text{rel}(f_i^v)$ is the relevance function, $\text{corr}(f_i^v, f_j^{v'})$ is the function to compute the correlation between pair of features, and $|S_a|$ is the size of the selected subset $S_a$ by ant $a$. In Eq. (6), the numerator of the fraction computes the average of the relevance values of features in the selected subset $S_a$, and the denominator calculates the average of the correlation values between each pair of distinct features in $S_a$.

*Probability Updating Rule*. The probability distribution is intended to make a view with the higher number of selected features more desirable for ants in the subsequent iterations. A new equation has been proposed and implemented to update probability distribution as follow:

$$p_v(t+1) = (1-\rho)p_v(t) + \rho\left[\frac{\text{count}(S^v)}{\text{count}(S)}\right], \quad (7)$$

where $p_v(t)$ and $p_v(t+1)$ indicate the probability values of view $v$ at time $t$ and $t+1$, respectively, and $S^v$ indicates the selected features subset in view $v$ during the current iteration.

## III. RESULTS AND DISCUSSION

We empirically evaluate the performance of the MAgentOmics as an unsupervised feature selection method in comparison to the mRMR-mv [17] which is a supervised multi-view feature selection method. The experiments are conducted upon ovarian cancer data derived from the TCGA.

The proposed method is implemented using Python and the code is publicly available on GitHub[2]. In the experiments, the classification accuracy of two widely used classifiers has been applied as the performance metric, logistic regression (LR) [26] and random forest (RF) [27]. The Scikit-learn library [28] was used to implement the presented classifiers.

To evaluate the average classification accuracy of the selected subsets, 5-fold cross-validation (CV) was used. Since k-fold CV may give rise to a noisy estimate of classifier performance, we repeated the CV procedure 5 times and reported the average values across all runs.

For the proposed method, a set of parameters is required to set, the maximum number of iterations $N_I = 30$, the number of ants in each view $N_A = 20$, $\alpha = \beta = 2$ to show equal importance of two heuristic information, the evaporation rate $\rho = 0.2$, initial pheromone level on each node $\tau_i^k(0) = 0.2, \forall k = 1, 2, \ldots, v$ , and state transition rules control parameter $q_0 = 0.7$.

For the mRMR-mv method, the absolute value of Pearson's correlation coefficient was considered as the redundancy function. Moreover, the mutual information implemented in Scikit-learn was selected as the relevance function according to the original paper [17].

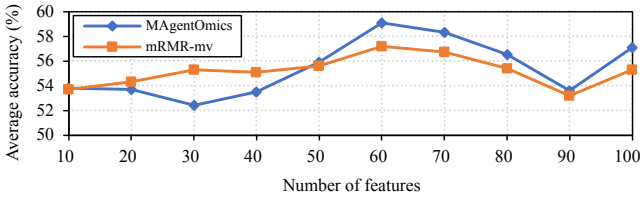
Fig. 2 illustrates the results of the comparison between the MAgentOmics method and the mRMR-mv method in terms of average classification accuracy (in %) of LR and RF classifiers over different selected features sizes. The reported results from Fig. 2a show that the proposed method has the better performance compared to mRMR-mv when the number of selected features is 10, 50, 60, 70, 80, 90, and 100. Additionally, Fig. 2b indicates that the classification accuracy of the proposed method is higher than the mRMR-mv method in most cases. For example, the classification accuracy of the MAgentOmics was 57.1% when 50 features were selected, while this value for mRMR-mv was 52.7%.

Overall, it can be observed that the best performance of the proposed method was 59.1% and 57.1% using LR and RF, respectively, while for mRMR-mv, these values were reported 57.2% and 55.3%, correspondingly.

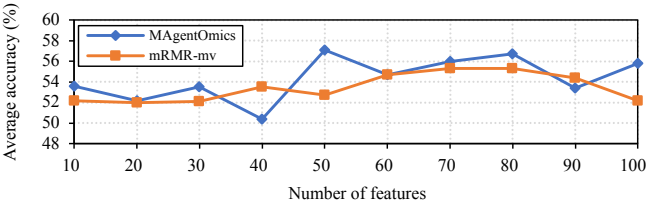
## IV. CONCLUSION AND FUTURE WORK

MAgentOmics is a multi-agent feature selection architecture designed for multi-omics data integration with the task of survival analysis. The architecture is based on the interactions of multiple ants which share their knowledge for finding the best global feature subset. In addition, we proposed a new fitness function to evaluate generated subsets by considering the maximum relevance and the minimum redundancy analyses. To assess the relative importance of

[2] https://github.com/SinaTabakhi/MAgentOmics.




(a) Average classification accuracy of logistic regression.




(b) Average classification accuracy of random forest.

Fig. 2. Average classification accuracy (in %), with respect to the number of selected features over 5-fold cross-validation.

each view in the feature selection process, we proposed a probability updating rule to iteratively update probability distribution by the concept of reinforcement learning. In other words, a view with a higher number of selected features is more likely to be chosen by ants in the following iterations.

The performance of the proposed method was compared to the supervised mRMR-mv feature selection method in terms of classification accuracy of the logistic regression and random forest classifiers. MAgentOmics achieved promising results on the TCGA ovarian cancer compared to the mRMR-mv method. Although the proposed fitness function analyzes the relevance and redundancy of the candidate feature subsets, it needs high computational time to calculate redundancies. Furthermore, it is an unsupervised function that may result in lower accuracy.

Future work will handle the design of new fitness functions to assess the subsets more accurately within lower computational time. Since the number of edges between pairs of features is high in the multi-omics data, considering a search space as a sparse graph can improve the execution time.

## References


[1] M. E. Barefoot, R. S. Varghese, Y. Zhou, C. D. Poto, A. Ferrarini, and H. W. Ressom, "Multi-omic pathway and network analysis to identify biomarkers for hepatocellular carcinoma," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2019, pp. 1350–1354, 2019.

[2] L. Tong, J. Mitchel, K. Chatlin, and M. D. Wang, "Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 225, 2020.

[3] H. R. Hassanzadeh, J. H. Phan, and M. D. Wang, "A semi-supervised method for predicting cancer survival using incomplete clinical data," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2015, pp. 210–213, 2015.

[4] M. Krassowski, V. Das, S. K. Sahu, and B. B. Misra, "State of the field in multi-omics research: from computational needs to data mining and sharing," *Frontiers in Genetics*, vol. 11, no. 1598, 2020.

[5] The Cancer Genome Atlas Research Network, "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," *Nature*, vol. 455, no. 7216, pp. 1061–1068, 2008.

[6] J. Zhang, J. Baran, A. Cros, J. M. Guberman, S. Haider, J. Hsu, Y. Liang, E. Rivkin, J. Wang, B. Whitty, M. Wong-Erasmus, L. Yao, and A. Kasprzyk, "International cancer genome consortium data portal–a one-stop shop for cancer genomics data," *Database*, vol. 2011, p. bar026, 2011.

[7] P. S. Reel, S. Reel, E. Pearson, E. Trucco, and E. Jefferson, "Using machine learning approaches for multi-omics data analysis: A review," *Biotechnology Advances*, vol. 49, p. 107739, 2021.

[8] M. Kang, E. Ko, and T. B. Mersha, "A roadmap for multi-omics data integration using deep learning," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab454, 2022.

[9] S. Tabakhi and P. Moradi, "Relevance–redundancy feature selection based on ant colony optimization," *Pattern Recognition*, vol. 48, no. 9, pp. 2798–2811, 2015.

[10] M. Picard, M.-P. Scott-Boyer, A. Bodein, O. Périn, and A. Droit, "Integration strategies of multi-omics data for machine learning analysis," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 3735–3746, 2021.

[11] L. C. Stetson, T. Pearl, Y. Chen, and J. S. Barnholtz-Sloan, "Computational identification of multi-omic correlates of anticancer therapeutic response," *BMC Genomics*, vol. 15, no. 7, p. S2, 2014.

[12] B. Tang, Z. Pan, K. Yin, and A. Khateeb, "Recent advances of deep learning in bioinformatics and computational biology," *Frontiers in Genetics*, vol. 10, 2019.

[13] T. Wang, W. Shao, Z. Huang, H. Tang, J. Zhang, Z. Ding, and K. Huang, "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nature Communications*, vol. 12, no. 1, p. 3445, 2021.

[14] Z. Huang, X. Zhan, S. Xiang, T. S. Johnson, B. Helm, C. Y. Yu, J. Zhang, P. Salama, M. Rizkalla, Z. Han, and K. Huang, "SALMON: survival analysis learning with multi-omics neural networks on breast cancer," *Frontiers in Genetics*, vol. 10, 2019.

[15] L. Cantini, P. Zakeri, C. Hernandez, A. Naldi, D. Thieffry, E. Remy, and A. Baudot, "Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer," *Nature Communications*, vol. 12, no. 1, p. 124, 2021.

[16] Y. Chen, M. E. Barefoot, R. S. Varghese, K. Wang, C. Di Poto, and H. W. Ressom, "Integrative analysis to identify race-associated metabolite biomarkers for hepatocellular carcinoma," in *42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2020, pp. 5300–5303, 2020.

[17] Y. El-Manzalawy, T.-Y. Hsieh, M. Shivakumar, D. Kim, and V. Honavar, "Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data," *BMC Medical Genomics*, vol. 11, no. 3, p. 71, 2018.

[18] S. Tabakhi, A. Najafi, R. Ranjbar, and P. Moradi, "Gene selection for microarray data classification using a novel ant colony optimization," *Neurocomputing*, vol. 168, pp. 1024–1036, 2015.

[19] W. Ma, X. Zhou, H. Zhu, L. Li, and L. Jiao, "A two-stage hybrid ant colony optimization for high-dimensional feature selection," *Pattern Recognition*, vol. 116, p. 107933, 2021.

[20] L. C. Sekhar and R. Vijayakumar, "Feature selection using ant colony optimization and weighted visibility graph," in *Evolution in Computational Intelligence*, pp. 17–32, Springer Singapore, 2021.

[21] M. Dorigo and T. Stützle, *Ant Colony Optimization: Overview and Recent Advances*, vol. 146 of *International Series in Operations Research & Management Science*, pp. 227–263. Springer US, 2010.

[22] X. Zhang, Y. Xing, K. Sun, and Y. Guo, "OmiEmbed: a unified multitask deep learning framework for multi-omics data," *Cancers*, vol. 13, no. 12, p. 3047, 2021.

[23] R. R. Vangimalla and J. Sreevalsan-Nair, "HCNM: heterogeneous correlation network model for multi-level integrative study of multi-omics data for cancer subtype prediction," *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2021, pp. 1880–1886, 2021.

[24] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier Science, 2008.

[25] T. M. Mitchell, *Machine Learning*. McGraw-Hill, Inc., 1997.

[26] S. Le Cessie and J. C. Van Houwelingen, "Ridge estimators in logistic regression," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 41, no. 1, pp. 191–201, 1992.

[27] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.