Novel feature selection method via kernel tensor decomposition for improved multi-omics data analysis

Y-h. Taguchi¹* and Turki Turki²

Abstract
Background: Feature selection of multi-omics data analysis remains challenging owing to the size of omics datasets, comprising approximately $10^2$–$10^5$ features. In particular, appropriate methods to weight individual omics datasets are unclear, and the approach adopted has substantial consequences for feature selection. In this study, we extended a recently proposed kernel tensor decomposition (KTD)-based unsupervised feature extraction (FE) method to integrate multi-omics datasets obtained from common samples in a weight-free manner.

Method: KTD-based unsupervised FE was reformatted as the collection of kernelized tensors sharing common samples, which was applied to synthetic and real datasets.

Results: The proposed advanced KTD-based unsupervised FE method showed comparative performance to that of the previously proposed KTD method, as well as tensor decomposition-based unsupervised FE, but required reduced memory and central processing unit time. Moreover, this advanced KTD method, specifically designed for multi-omics analysis, attributes $P$ values to features, which is rare for existing multi-omics–oriented methods.

Conclusions: The sample R code is available at https://github.com/tagtag/MultiR/.

Keywords: Tensor decomposition, Feature selection, Multiomcis, Kernel trick

Background
Feature selection with multi-omics datasets has been a long-standing challenge for bioinformatics. Among the numerous proposed methods adapted to multi-omics data analysis [1, 2], only few are capable of performing feature selection. Most of these methods fail to implement feature selection because multi-omics data analysis has a strong tendency to involve a small number ($=n$) of samples with a large number ($=p$) of features, commonly referred to as the large $p$ small $n$ problem [3], posing difficulty for accurate feature selection. Features should have a sufficiently small $P$ value to be selected under the null hypothesis. Since the raw $P$ values must be heavily corrected for multiple comparisons when dealing with multi-omics datasets, $P$ values become larger and inevitably less significant; thus, attributing significant $P$ values to individual features, even after correction, is difficult. However, since the number of samples (i.e., conditions) is less than that of features (i.e., variables), labels or values attributed to samples can be accurately predicted by any model (when the number of conditions is less than the number of variables, either the labels or values attributed to samples may be predicted, even if the variables are purely random numbers).

In multi-omics analysis, it is difficult to obtain large sample sizes since multiple observations, each of which corresponds to individual omics approaches, must be performed. In this sense, the required cost and time

*Correspondence: tag@granular.com
¹ Department of Physics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
Full list of author information is available at the end of the article

© The Author(s) 2022. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
are multiplied in proportion to the number of omics approaches considered. This often results in a smaller number of samples to which multi-omics measurements are performed when only limited experimental resources are available.

Principal component analysis (PCA) and tensor decomposition (TD)-based unsupervised feature extraction (FE) [4] have been proposed to be applied to feature selection for addressing the large p small n problem. Thus, these approaches are also suitable for feature selection in multi-omics analysis. Recently, the TD-based method was extended to kernel TD (KTD)-based unsupervised FE [5], which was applied to integrated analysis of N⁶-methyladenosine (m⁶A) and gene expression data [6]. These methods attribute P values to features, which is critical since this enables evaluation of the significance of the selected features, which is rarely possible using other methods applicable to multi-omics datasets [1, 2]. In spite of this advantage, there are limitations of PCA, TD, and KTD-based unsupervised FE when applied to multi-omics data analysis. PCA is inferior to TD when aiming for an integrated analysis of multiomics data. PCA failed to identify genes whose expression and methylation levels are altered simultaneously, but TD could [7]. Although KTD and TD successfully integrated two omics data, they could not achieve the following two points,

1. Reduction of required memory and long CPU time
2. Integration of more than two types of omics data simultaneously (see Discussions below). We here described a modification of the KTD-based unsupervised FE method to be more suitable for multi-omics data analysis. Although only a small modification was implemented, it nevertheless resulted in more flexibility for multi-omics data analysis, which was verified using synthetic and real data.

Before explaining the results, we briefly discuss the relationship between feature selection and feature extraction, both of which are employed when we are forced to deal with the large p small n problem. The former, feature selection, is more straightforward than the latter; it reduces the number of features less than the number of samples. On the other hand, feature extraction is more indirect, since it generates a limited number of new features from the original large number of features. In this study, we have employed a mixed strategy of these two. We first generated new features using feature extraction and selected features using the generated features.

### Methods

#### Extended KTD-based unsupervised FE method

Suppose that we have K multi-omics datasets with N_k features formatted as tensors sharing sample indices j_1, \ldots, j_m as:

\[
\mathbf{x}_{i; j_1, j_2, \ldots, j_m} \in \mathbb{R}^{N_k \times M_1 \times M_2 \times \cdots \times M_m},
\]

\( j_s, (1 \leq j_s \leq M_s) \) refers to the j_sth measurement in the sth experimental type. \( M_s, (1 \leq s \leq m) \) is the number of measurements in the sth experimental type. Typical examples of m experimental conditions include human subjects, tissues, and time points. For example, if the measurements are performed for \( M_2 \) tissue types from \( M_1 \) individuals at \( M_3 \) time points, the total number of samples is \( M_1 \times M_2 \times M_3 \). \( \mathbf{x}_{i; j_1, j_2, \ldots, j_m} \) refers to the i\textsuperscript{th} feature of the k\textsuperscript{th} omics dataset. When K types of omics data are measured for each sample, \( k \in [1, K] \).

\( \mathbf{x}_{i; j_1, j_2, \ldots, j_m} \) can be kernelized as:

\[
\mathbf{x}_{k; j_1, j_2, \ldots, j_m} = K^k(\mathbf{x}_{i; j_1, j_2, \ldots, j_m}, \mathbf{x}_{i; j_1, j_2, \ldots, j_m})
\]

\( \in \mathbb{R}^{K \times M_1 \times M_2 \times \cdots \times M_m} \),

where \( K^k \) is an arbitrary kernel applied to \( \mathbf{x}_{i; j_1, j_2, \ldots, j_m} \).

Higher-order singular-value decomposition (HOSVD) [4] is then applied to \( \mathbf{x}_{k; j_1, j_2, \ldots, j_m} \), resulting in Eq. (4),

\[
\mathbf{x}_{k; j_1, j_2, \ldots, j_m} = \sum_{\ell_1=1}^{M_1} \cdots \sum_{\ell_m=1}^{M_m} \sum_{\ell_{m+1}=1}^{M_{m+1}} \cdots \sum_{\ell_K=1}^{M_K} \sum_{s=1}^{K} G(\ell_1, \ldots, \ell_m, \ell_{m+1}, \ldots, \ell_{2m+1}, \ell_{2m+1}) \times \left( \prod_{s=1}^{m} U_{\ell_s, j_{s}} \right) \left( \prod_{s'=1}^{m} U_{\ell_{s+m+1}, j_{s'}} \right) U_{\ell_{2m+1}, k},
\]

where \( \ell_s, (1 \leq s \leq 2m) \) refers to the \( \ell_s \text{th singular-value vector attributed to the } s \text{th experimental type for } 1 \leq s \leq m \text{ and the } (s-2m) \text{th experimental type for } m+1 \leq s \leq 2m, \text{ respectively. } \ell_{2m+1} \text{ refers to the } \ell_{2m+1} \text{th singular-value vector attributed to the omics datasets, } k. \)

\( U_{\ell_s, j_s} \in \mathbb{R}^{M_s \times M_s} \) and \( U_{\ell_{2m+1}, k} \in \mathbb{R}^{K \times K} \) are singular-value matrices, which are also orthogonal matrices, \( \sum_{s=1}^{M_s} U_{\ell_s, j_s} U_{\ell_s, j_s} = \delta_{\ell_s, \ell_s}, \sum_{s=1}^{M_{m+1}} U_{\ell_{s+m+1}, j_{s+m+1}} U_{\ell_{s+m+1}, j_{s+m+1}} = \delta_{\ell_s, \ell_s}, \sum_{s=1}^{K} U_{\ell_{2m+1}, k} U_{\ell_{2m+1}, k} = \delta_{kk} \)

where \( \delta_{\ell_s, \ell_s}, \delta_{j_s, j_s}, \delta_{j_{s+m+1}, j_{s+m+1}}, \delta_{kk} \) are Kronecker’s delta. Because of symmetry, \( U_{\ell_{2m+1}, k} = U_{\ell_{2m+1}, k}^T \).

\( G(\ell_1, \ldots, \ell_m, \ell_{m+1}, \ldots, \ell_{2m+1}) \in \mathbb{R}^{M_1 \times \cdots \times M_m \times M_{m+1} \times \cdots \times M_{2m+1} \times K} \) is a core tensor that represents the weight of individual terms composed of the products of singular-value vectors. Here, one should note that \( U_{\ell_{2m+1}, k} \) represents the balance (weight) between multi-omics datasets, which
usually must be pre-defined manually in the case of a conventional supervised learning approach for FE.

Next, the \( u_{ij} \) values that are of interest from the biological point of view (e.g., distinct values between the two classes being compared) must be identified. Using these selected \( u_{ij} \) values, singular-value vectors are derived and assigned to \( k \) as

\[
u_k = \sum_{j=1}^{M} \cdots \sum_{j_m=1}^{M_m} x_{ijk_1 \cdots jm} \left( \prod_{s=1}^{m} u_{ls} \right),
\]

where \( \ell = (\ell_1, \ldots, \ell_m) \).

Finally, \( P \) values are attributed to \( k \) assuming that the \( u_k \) values obey a multivariate Gaussian distribution (null hypothesis) as

\[
P_k = P_{\chi^2} \left[ \sum_{\ell} \left( \frac{u_{\ell k}}{\alpha_\ell} \right)^2 \right],
\]

where \( P_{\chi^2}[> x] \) is a cumulative \( \chi^2 \) distribution in which the argument is larger than \( x \) and \( \alpha_\ell \) is the standard deviation. Here, summation is taken over \( \ell \) selected as being of interest. \( P \) values are then computed by the \texttt{pchisq} function in R [8].

The obtained \( P_k \) values are corrected using the Benjamini–Hochberg (BH) criterion [4] and the \( k \) values associated with the adjusted \( P \) values less than the established threshold (typically 0.01) are selected. Correction by the BH criterion is performed by the \texttt{p.adjust} function in R with the option of method="BH".

**Synthetic dataset**

A synthetic dataset was derived in the form of a tensor, \( x_{ijk} \in \mathbb{R}^{N \times M \times K} \), as

\[
x_{ijk} = \begin{cases} \epsilon_{ijk} + a_j i \leq N_1, kN_1 < i \leq (k+1)N_1 \\ \epsilon_{ijk} \\ \end{cases},
\]

where

\[
a_j = 1 + \frac{M - 1}{j}
\]

and \( \epsilon_{ijk} \in (0, 1) \) is a uniform random real number that emulates the residuals. \( N \) is the number of variables, \( j \) is a variable that adds order to the \( j \)th sample, and \( k \) is the \( k \)th omics data. This synthetic dataset assumes that only the top \( N_1 \) features among \( N \) features have dependency on \( i \) independent of \( k \). \( x_{ijk}, (kN_1 < i \leq (k+1)N_1) \) also has dependence on \( j \), but in an omics(\( k \))-dependent manner. The dependence on \( j \) is a linear increase upon \( j \).

One hundred ensembles of \( x_{ijk} \) were generated and the performances were averaged. First, a linear kernel was generated as

\[
x_{kij} = K^k(x_{ij}, x_{j^\prime k}) = \sum_{i=1}^{N} x_{ijk} x_{ij^\prime k} \in \mathbb{R}^{K \times M \times M}.
\]

Next, HOSVD was applied, resulting in

\[
x_{kij} = \sum_{\ell_1=1}^{M} \sum_{\ell_2=1}^{M} \sum_{\ell_3=1}^{K} G(\ell_1, \ell_2, \ell_3) u_{\ell_1 k} u_{\ell_2 j} u_{\ell_3 k}.
\]

(10)

(for the dimensions of the datasets, \( N, M, \) and \( K \), see the legend to Table 1). Since it was observed that \( u_{2j} \) always had the largest correlation with \( a_j \) and \( u_{1k} \) was always constant, regardless of \( k \), \( u_{2j} \) and \( u_{1k} \) were employed to compute

\[
u_{\ell_1 \ell_2} = \sum_{j=1}^{M} \sum_{k=1}^{3} x_{ijk} u_{\ell_1 j} u_{\ell_2 k}
\]

(11)

for attributing the \( P \) values

\[
P_k = P_{\chi^2} \left[ \left( \frac{u_{\ell_1 \ell_2}}{\alpha_{\ell_1 \ell_2}} \right)^2 \right]
\]

(12)

with \( \ell_1 = 2 \) and \( \ell_2 = 1 \). The \( P_k \) values were corrected using the BH criterion and the \( i \) values associated with adjusted \( P_k \) less than 0.01 were selected.

To demonstrate the difficulty of this task, linear regression was used as an alternative method

\[
x_{ijk} = \alpha_{ik} a_j + \beta_{ik},
\]

(13)

where \( \alpha_{ik} \) and \( \beta_{ik} \) are regression coefficients, and \( a_j \) is defined in Eq. (8). Since \( x_{ijk} \) s with distinct \( k \) differ, distinct models were applied to each. After computing BH-corrected \( P \) values, the \( i \) associated with adjusted \( P \) values less than 0.01 were selected.

When the least absolute shrinkage and selection operator (lasso) [9] was applied, the maximum number of features selected were considered, although lasso can select at most as many as \( M \) features, which is less than the number of features coincident with \( a_j \), \( 2N_1 \).

When random forest (rf) [10] was applied to synthetic data sets, there were two ways to select features. First, features with nonzero importance, which is an evaluation measure provided by rf, were selected. Next, in order

| \( i \leq N_1 \) | \( i > N_1 \) |
|---|---|
| Adjusted \( P_i \leq 0.01 \) | 7.06 | 2.94 |
| Adjusted \( P_i > 0.01 \) | 0.04 | 989.96 |
to reduce the number of selected features, the top most $2N_j$ features having a larger absolute importance were selected.

**Multi-omics hepatitis B virus (HBV) vaccine dataset**

The real multi-omics HPV vaccine datasets were based on 75 samples measured in 15 individuals at five subsequent time points (i.e., 0, 1, 3, 7, and 14 days) after HBV vaccine treatment. Gene expression was measured using RNA-sequencing technology and methylation profiles were measured using microarray technology. The proteome was measured for whole blood cells (WBCs) as well as plasma. Since this multi-omics dataset is composed of four types of omics data measured for 15 individuals at five time points, it is thus formatted as a tensor.

Gene expression and methylation profiles were retrieved from the gene expression omnibus (GEO) database using the GSE155198 and GSE161020 datasets, respectively. For gene expression profiles, the GSE155198_RAW.tar dataset was available in the Supplementary File section of GEO. Individual files that included count number of mapped reads toward genes were collected and integrated as a single file. Individual files are named according to the format “GSMXXXXXX_GRnm_Vm.count.txt.gz”, where XXXXXX, nn, and m are integers; $nn \in \{01, 02, 03, 04, 05, 06, 07, 10, 11, 13, 15, 17, 18, 19\}$ identifies the 15 individuals; and $m \in \{3, 4, 5, 6, 7\}$ identifies the four time points. Files were loaded into R as a data frame using the `read.csv` command. Data frames were bound into a single data frame using the `cbind` command in R. For the methylation profiles, the GSE161020_series_matrix.txt.gz dataset, available in the GEO Series Matrix File(s) section, was used as is. The file was loaded into R as a data frame using the `read.csv` command. The first column of the data frame is an identifier in the form of cgYYYYYYYY, where YYYYYYYY is an integer. Since the methylation profile was measured with microarray technology, the identifier can be annotated with the reference to the microarray annotation file, GPL6480-9577.txt.gz, which is available under GEO ID GPL6480. Since the other 75 column names are in the form of GSMXXXXXX, the columns were reordered with reference to the columns of the data frame generated from the gene expression profiles as described above. The proteome dataset was obtained from ProteomeXchange [11] using ID PXD020474. Two files (GR01,04,09,10,11,13,15,17,18,19.txt and GR02,03,05,06,07.txt) were downloaded and loaded as data frames into R using the `read.csv` command. The fourth column of the data frame includes the protein IDs, which were used as identifiers for subsequent analysis. These two data frames were merged with the row names of the union of the protein identifier. Missing observations were filled with zeros. Since the first and second rows have the format “GRnm” and “Visit mi”, respectively, the columns can also be reordered with reference to the column names of the data frame generated by gene expression profiles as described above.

$x_{i,j_{12}} \in \mathbb{R}^{N_{x} \times 3 \times 15}$; that is, the number of values of the number of $i_k$ features of the $k$th feature type measured at the $j_{1}$th time point for $j_2$ individuals. These values are standardized as $\sum_{j_{1}=1}^{N_{x}} x_{i,j_{12}} = 0$ and $\sum_{j_{1}=1}^{N_{x}} x_{i,j_{12}}^2 = N_{x}$; $N_{x} = 687582$ $(k = 1$ : methylation), 35829 $(k = 2$ : gene expression), 1588 $(k = 3$ : WBC proteome and $k = 4$ : plasma proteome). Figure 1 schematically illustrates the analysis method for the HBV vaccination datasets.

A linear kernel was employed as follows:

$$K^k (x_{i,j_{12}}, x_{i',j'_{12}}) = \sum_{i_k=1}^{N_k} x_{i,j_{12}} x_{i',j'_{12}},$$

Then, a tensor was added:

$$x_{k,j_{12}} \in \mathbb{R}^{4 \times 3 \times 15 \times 15 \times 15},$$

where $4$ stands for four omics data, $5$ stands for five time points and $15$ stands for fifteen individuals. Applying HOSVD to $x_{k,j_{12}}$ results in

$$x_{k,j_{12}} = G_{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5} u_{\ell_1} u_{\ell_2} u_{\ell_3} u_{\ell_4} u_{\ell_5},$$

where $G \in \mathbb{R}^{3 \times 15 \times 15 \times 15}$, $u_{\ell_1}, u_{\ell_2} \in \mathbb{R}^{3 \times 5}$, $u_{\ell_3}, u_{\ell_4} \in \mathbb{R}^{15 \times 5}$, and $u_{\ell_5} \in \mathbb{R}^{4 \times 4}$. Since $x_{k,j_{12}} = x_{k,j_{12}}$ and $x_{k,j_{12}} = x_{k,j_{12}}$ because of symmetry, $G_{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5} = G_{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5}$, $\{u_{\ell_1}\} = \{u_{\ell_1}\}$, and $\{u_{\ell_2}\} = \{u_{\ell_2}\}$.

The singular-value vectors of interest were as follows:

- $u_{\ell_1}$ and $u_{\ell_2}$ should be significantly dependent on time points corresponding to $j_1$ and $j_2$.
- $u_{\ell_3}$ and $u_{\ell_4}$ should be independent of individuals corresponding to $j_2$ and $j_3$.
- $u_{\ell_5}$ should be common between distinct omics measurements.

As a result, $\ell_1 = \ell_3 = 2$, $\ell_2 = \ell_4 = 1$, $\ell_5 = 1$ satisfies the required conditions (Fig. 2). Then,

$$u_{21j_{12}} = \sum_{j_{1}=1}^{15} x_{i,j_{12}} u_{21j_{12}},$$

$$u_{21j_{12}} = \sum_{j_{1}=1}^{15} x_{i,j_{12}} u_{21j_{12}}.$$
The computed \( P_{ik} \) values were corrected using the BH criterion and \( k \) is associated with either \( P_{ik} < 0.01 \) (for gene expression and methylation) or with \( P_{ik} < 0.05 \) (for the two proteomes) were successfully selected (the full lists of selected features are available in Additional file 2, 3, 4, 5: Data S1–S4).

**Kidney cancer multi-omics datasets**

Full description of the compilation of the kidney cancer multi-omics datasets is available in the related study [12]. In brief, there were two sets of multi-omics kidney cancer data, each of which was composed of messenger RNA (mRNA) and microRNA (miRNA) expression profiles.
The first dataset was obtained from The Cancer Genome Atlas (TCGA) and included 253 kidney tumors and 71 normal kidneys. The second dataset was obtained from GEO (GSE16441), and included 17 patients and 17 healthy controls. The method by which these two datasets were preprocessed is described in [12]. The dataset was formatted as \( x_{ikj} \in \mathbb{R}^{N_k \times M} \), which represents the \( i_k \) expression levels of \( j \) subjects (\( k = 1 \) for mRNA and \( k = 2 \) for miRNA).

A linear kernel was employed as

\[
K^k(x_{ikj}, x_{ik'j'}) = \sum_{i_k=1}^{N_k} x_{ikj} x_{ik'j'} \in \mathbb{R}^{K \times M \times M}. \tag{19}
\]

For the first dataset (i.e., TCGA data), \( M = 324 \), whereas for the second dataset (i.e., GEO data), \( M = 37 \). Then, a tensor was added

\[
x_{kj'} = K^k(x_{ikj}, x_{ik'j'}) \in \mathbb{R}^{2 \times M \times M}. \tag{20}
\]

Applying HOSVD to \( x_{kj'} \) results in

\[
x_{kj'} = \sum_{k=1}^{2} G(\ell_{1}, \ell_{2}, \ell_{3}) u_{\ell_{1}j} u_{\ell_{2}j'} u_{\ell_{3}k}, \tag{21}
\]

where \( G \in \mathbb{R}^{M \times 2}, \ u_{\ell_{1}j}, u_{\ell_{2}j'} \in \mathbb{R}^{M \times M} \), and \( u_{\ell_{3}k} \in \mathbb{R}^{2 \times 2} \). Since \( x_{kj'} = x_{kj'} \) because of symmetry, \( G(\ell_{1}, \ell_{2}, \ell_{3}) = G(\ell_{1}, \ell_{3}, \ell_{2}) \). Then, singular-value vectors were identified such that \( u_{\ell_{1}j} \) and \( u_{\ell_{2}j'} \) were significantly distinct between healthy controls and patients. As a result, \( \ell_{1} = \ell_{2} = 2 \) satisfies the required conditions (Fig. 3). Then,

\[
u_{2jk} = \sum_{j=1}^{M} x_{ikj} w_{2j} \tag{22}
\]

is computed and

\[
P_{i_k} = P_{X^2} \left( \frac{u_{2ik}}{\sigma_2} \right)^2. \tag{23}
\]

The computed \( P_{i_k} \) values were corrected using the BH criterion and \( i_k \)s associated with adjusted were successfully selected.

**Results**

**Synthetic dataset**

Table 1 shows the confusion matrix obtained by the proposed KTD-based unsupervised FE. Among the 10 features associated with \( a_i \), approximately seven features were correctly selected, whereas false positives were almost zero. Thus, KTD-based unsupervised FE successfully selected features correlated with \( a_i \).

Table 2 shows the confusion matrix obtained by linear regression-based FE. Essentially, no features correlated with \( a_i \) were selected. Thus, regression analysis did not select any features correlated with \( a_i \).

|                | Linear regression | Lasso | RF |
|----------------|-------------------|-------|----|
|                | Adjusted          |        |     |
| \( k = 1 \)    |                   |       |     |
| \( i \leq 2N_k \) | 0.07              | 19.93 |     |
| \( i > 2N_k \)  | 0.03              | 979.97|     |
| \( k = 2 \)    |                   |       |     |
| \( i \leq N_k 2N_k < i \leq 3N_k \) | 0.07 | 19.93 |     |
| \( i \leq 3N_k \) | 0.07              | 19.93 |     |
| Other than above | 0.01              | 979.99|     |
| \( k = 3 \)    |                   |       |     |
| \( i \leq N_k 3N_k < i \leq 4N_k \) | 0.09 | 19.91 |     |
| Other than above | 0.01              | 979.99|     |

For cases when rf was employed, the results when the top most \( 2N_k \) features with larger absolute importance were selected have also been shown in parentheses.
These results demonstrated that an apparently simple and easy problem became difficult when it is a large $p$ small $n$ problem, whereas KTD-based unsupervised FE was able to handle this problem to some extent. These advantages have also been observed in PCA- and TD-based unsupervised FE [4].

Although the methods that did not attribute $P$ values to features were not of interest, since the capability of attributing $P$ values to features is a great advantage of KTD-based unsupervised FE, as emphasized in the Background, lasso was employed as another method for comparison. Although lasso regression does not attribute $P$ values to features, the model was fitted to feature selection in a large $p$ small $n$ problem to demonstrate the difficulty of feature selection in the synthetic dataset. Table 2 shows the confusion matrix, which is clearly inferior to that shown in Table 1. Although the KTD-based unsupervised FE approach correctly selected at least 7 out of 10 features (70%), which was correlated with $a_i$ with essentially no false positives, lasso selected at most 5 out of 20 features (only 25%), which was correlated with $a_i$ with two false positives (approximately half of the true positives). In addition to lasso, we also tested rf as an alternative method that cannot attribute $P$ values (Table 2). First, we have selected features with nonzero importance; although most of (ca. 17) features among the 20 features coincident with $a_i$ are selected, almost half of features not coincident with $a_i$ are also wrongly selected. Thus, rf is clearly inferior to KTD-based unsupervised FE. One might wonder if the top 20 features with a larger absolute importance are selected. Only five out of 20 features are selected. Thus, rf is still inferior to TD-based unsupervised FE even if limited and most important features are selected. This suggested that even if methods that could not attribute $P$ values to features can be considered, they would not outperform the KTD-based unsupervised FE method.

Thus, subsequently, we only focused only methods that attributed $P$ values to features.

HBV vaccine dataset
To validate genes selected by KTD-based unsupervised FE, the selected genes were uploaded to Enrichr [13]. Initially, 1335 genes associated with 2077 methylation probes selected by KTD-based unsupervised FE were uploaded. Many transcription factors (TFs) were significantly predicted to target these 1335 genes (Table 3). These 21 TFs were then uploaded to Enrichr again; Additional file 1: Table S1 shows the top 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the “KEGG 2019 HUMAN” Enrichr category (the full list is available in Additional file 2: Data S1). It was clear that these data included many biologically reasonable KEGG pathways (for details, see the Discussion section below).

Eight genes associated with 11 probes were identified as differentially expressed genes (DEGs) when gene expression profiles were considered (Table 4), which were uploaded to Enrichr. Conversion of probe names to gene symbols was performed by the ID converter tool of DAVID [14]. Additional file 1: Table S2 shows the top 10 GEO profiles enriched in the Enrichr “Disease Perturbations from GEO up/down” category (the full list is available in Additional file 3: Data S2). Although many neurodegenerative profiles not apparently related to the HBV vaccine are listed, the rationalization for their inclusion is provided in the Discussion section.

Finally, two sets of proteins identified as DEGs in WBC and plasma profiles (Table 4) were uploaded to Enrichr. Additional file 1: Tables S3 and S4 show the top 10 enriched GEO profiles identified in the Enrichr “Disease Perturbations from GEO up/down” category when proteins for the WBC and plasma sections in Table 5 were uploaded to Enrichr (the full list is available in Additional file 3, 4: Data S3 and S4). Other than the enriched

---

**Table 3**

| TFs          | ZNF217, TCF4, STAT3, SMARCD1, WT1, FOXA2, PAX3-FKHR, SMAD4, SMAD3, SOX9, TFAP2C, YAP1, AR, SOX2, CTNNB1, VDR, PIAS1, TEAD4, MITF, HNF4A, SUZ12 |
|--------------|----------------------------------------------------------------------------------------------------------------------------------|

| TFs enriched in the “ChEA 2016” Enrichr category (adjusted $P$ values < 0.05) when 1335 genes associated with 2077 methylation probes selected by KTD-based unsupervised FE were considered (the full list is available in Additional file 2: Data S1) |

**Table 4**

| Gene symbols | S100A9, CD74, hba1, ACTB, HBB, HBA2, MALAT1, COX1 |
|--------------|------------------------------------------------------------------|
| WBC          | HIST1H2B1, HIST2H2F, HIST1H2B1, HIST1H2B2, HIST1H2BD, ACTG1, HIST1H2B1, HIST1H2BD, PFN1, HIST1H2B1, HIST3H2B8, ACTB, HBB, HBA2, HIST1H2A, HIST1H2B1, HIST1H2BC, HIST1H2BO, HIST2H2E, HIST1H2B, HIST1H2BH, HIST1H2BD |
| Plasma       | FGA, HP, GSN, ALB, FGG, IGL5S, APOA1, SERPINA1, ORM1, TF, GC, CP, C4A, CSF3R, A2M, HPX, HRG, A1BG, CFH, APOB, C3, CLEC14A |

Eight genes associated with 11 probes identified as DEGs when gene expression profiles were considered. Proteins identified as DEGs when gene expression profiles in the proteome were considered.
reasonable hepatitis-related GEO profiles, some neurodegenerative disease-related GEO profiles were also enriched, as shown in Additional file 1: Table S3, which are further explored in the Discussion section.

Since there are many enriched biological processes and pathways listed in Additional file 1: Tables S1–S4, the genes and proteins selected in this section may not be artifacts, but likely have a true biological basis.

### Kidney cancer

hsa-mir-200c and hsa-mir-141 were selected from TCGA, and hsa-miR-141, hsa-miR-210, and hsa-miR-200c were selected from GEO. Thus, these miRNAs are highly coincident with each other, even more so than reported in previous works [5, 12] where TD- as well as KTD-based unsupervised FE was applied to TCGA and GEO datasets. For mRNA, there were five common genes selected between the TCGA and GEO datasets (Table 6, \( P = 6.7 \times 10^{-5}, \text{Odds ratio: 13.13} \)).

### Discussion

There are several advantages in the proposed implementation of KTD-based unsupervised FE compared with previously proposed versions of KTD-based unsupervised FE [5], as well as TD-based unsupervised FE [4] in the context of application to the integration of multi-omics datasets. For example,

1. The present implementation can reduce the required computational memory. As a result, required computational time can be reduced as well.
2. The present implementation can integrate more than two omics data in a straight manner.

Although the point was achieved in KTD-based unsupervised FE [5], too, the primary advantage of the proposed method is to achieve the above two simultaneously, as discussed below. As for point 1, when the original implementation of TD-based unsupervised FE [4] is applied to the integration of multi-omics data, \( (e.g., x_{ij} \in \mathbb{R}^{N \times M} \) and \( x_{hj} \in \mathbb{R}^{H \times M}, \) which corresponds to the \( i \)-th or \( h \)-th omics data of the \( j \)-th sample), HOSVD is applied to

\[
x_{ih} = \sum_{j} x_{ij}x_{hj} \in \mathbb{R}^{N \times H},
\]

Since \( H, N \gg M, \) this was not an effective implementation. When KTD-based unsupervised FE [5] is applied, HOSVD is applied to

\[
x_{ij''} = \sum_{j'} \left( \sum_{i} x_{ij}x_{ij'} \right) \left( \sum_{h} x_{hj'}x_{hj''} \right) \in \mathbb{R}^{M \times M}.
\]

This drastically reduced the required memory and central processing unit (CPU) time. As for point 2, nevertheless, it was unclear how more than two omics datasets could be integrated. In the implementation introduced in this paper,

\[
x_{kij'} = \sum_{i} x_{ijk}x_{ij'k} \in \mathbb{R}^{K \times M \times M},
\]

which corresponds to the \( i \)-th measurement of the \( k \)-th omics data of the \( j \)-th sample, where \( K \) is the total number of omics datasets. Since HOSVD is applied to \( x_{kij'}, \) any number of multi-omics datasets may be handled. This slight modification drastically increased the ability of KTD-based unsupervised FE to handle multi-omics datasets. Thus, it is obvious that the present implementation has at least one advantage over the past KTD implementations.

Kernels \( K^k \) were highly correlated between mRNA \((k = 1)\) and miRNA \((k = 2)\) for the kidney cancer data [15] (Fig. 4). Kernels \( K^k \) for gene expression profiles and the proteome were also highly correlated when HBV vaccine experiments were considered (Table 7). Thus, it was obvious that the current formalism was very effective in identifying the coincidence between individual omics (in this case, mRNA, miRNA, and the proteome).

Although the list of enriched pathways in Additional file 1: Table S1 did not seem to be related to the HBV vaccine directly, there were indirect reasonable relationships. For example, the Hippo signaling pathway has recently

---

**Table 5** Proteins identified as DEGs when gene expression profiles in the proteome were considered

| WBC       | HIST1H2BJ, HIST1H2BF, HIST1H2BG, HIST1H2BB, HIST1H2BD, ACTG1, HIST1H2BL, HIST1H2BN, PFN1, HIST1H2BK, HIST3H2BB, ACTB, HBB, HBA2, HIST1H2BA, HIST1H2B, HIST1H2BC, HIST1H2BO, HIST2H2BE, HIST1H2BM, HBA1, HIST1H2BF, HIST1H2BE, HIST1H2BH |
| Plasma    | FGA, HP, GSN, ALB, FGG, IGLL5, APOA1, SERPINA1, ORM1, TF, GC, CP, C4A, CSF3R, A2M, HPX, HRG, A1BG, CFH, APOB, C3, CLEC14A |

**Table 6** Confusion matrix of selected mRNAs between TCGA and GEO datasets

| GEO | \( P > 0.01 \) | \( P < 0.01 \) |
|-----|----------------|----------------|
| TCGA| 17269          | 101            |
|     | 65             | 5              |

\( P = 6.7 \times 10^{-5}, \text{Odds ratio: 13.13} \)
been identified to be related to the immune system [16]. “Hepatitis B” was also significantly enriched (adjusted \( P \) value of \( 7.79 \times 10^{-3} \)), but it did not rank within the top 10 KEGG pathways in Additional file 1: Table S1. In addition, since patients with diabetes have a higher risk of HBV infection [17], it is reasonable that the two KEGG pathways “AGE-RAGE signaling pathway in diabetic complications” and “Maturity onset of diabetes in the young” were enriched. Although multiple cancer types were also enriched in these data, as shown in Additional file 1: Table S1, many cancer types other than liver cancer are known to be related to the risk of HBV infection [18].

Although genes identified as DEGs in relation to HBV vaccination were also enriched in various neurodegenerative diseases other than hepatitis (Additional file 1: Tables S2 and S3), this is a reasonable finding because viral hepatitis was reported to be related to Parkinson’s disease [19]. There are also known associations between hepatic functions and plasma amyloid-\( \beta \) levels [20]; cirrhosis patients with HBV infection have higher plasma A\( \beta \)40 and A\( \beta \)42 levels than patients with HBV-negative cirrhosis. More directly, Ji et al. [21] reported that the hepatitis B core VLP-based mis-disordered tau vaccine alleviated cognitive deficits and neuropathology progression in a Tau.P301S mouse model of Alzheimer’s disease. Thus, enrichment of neurodegenerative disease-related genes among the identified DEGs does not appear to be an artifact, but rather provides possible supportive evidence that KTD-based unsupervised FE detected side effects caused by vaccinations.

Other conventional univariate tools such as limma [22] and sam [23] cannot be used for these tasks since they are designed to handle categorical classes and thus cannot be applied to HBV vaccination data, which are only associated with time points and are not categorical. Although regression analysis was attempted for the synthetic dataset, there were no features correlated with dates. Thus, there were no univariate feature selection methods applicable to the HBV vaccination data that could identify features correlated with date. For the kidney cancer datasets, it has been extensively demonstrated that these conventional univariate tools such as limma [22] and sam [23] cannot compete with the TD-based unsupervised FE approach [5, 12]. Thus, no univariate feature selection method was identified that was superior to TD-based unsupervised FE when applied to kidney cancer datasets.

The performance of the proposed method has not been compared to other existing multi-omics–oriented methods [1, 2] because no suitable methods were identified for suitable comparison. First, most of the recently proposed cutting-edge methods adapted to multi-omics analysis are specific to a high-throughput sequencing (HTS) architecture. For example, MKpLMM [24]

| Table 7 Correlation between HBV vaccine experiment kernels. Upper triangle: Pearson correlation, lower triangle: \( P \) values |
|---|---|---|
| Gene expression | WBC | Plasma |
| Gene expression | – | 0.1279405 | 0.2192163 |
| WBC | \( 1.34 \times 10^{-11} \) | – | 0.4384998 |
| Plasma | \( 1.52 \times 10^{-31} \) | \( 8.98 \times 10^{-131} \) | – |

![Fig. 4 Scatter plot of kernels \( k^4 \) between messenger RNA (mRNA) and microRNA (miRNA). Upper: TCGA (Pearson correlation coefficient = 0.627, \( P = 0.00 \)), Lower: GEO (Pearson correlation coefficient = 0.349, \( P = 2.32 \times 10^{-34} \))](image-url)
requires genomic coordinates, which are not available for the datasets analyzed in this study. Similarly, csaw [25] requires a bed file, which is also not available for the present dataset. Since the purpose of the present study was to propose a more flexible method that is not specific to the HTS architecture and the datasets employed in this study were not obtained using HTS, these were considered unsuitable methods for comparison with the proposed implementation of KTD-based unsupervised FE. Second, other methods that are not specific to HTS lack the statistical validation of feature selection (i.e., no ability to attribute \( P \) values to features). For example, although MOFA [26] is not specific to HTS, it does not have the ability to select features; thus, we were not able to compare its performance with that of our proposed method. Although DIABLO [27] is also not specific to the HTS architecture and has feature selection ability, there is no functionality to attribute \( P \) values to individual features; thus, features cannot be selected based on statistical significance, and therefore, DIABLO was considered to be outside of the scope of this study. FSMKL [28], which is also not specific to the HTS architecture, does have the ability to add statistical scores to individual features; however, selecting features based on statistical scores is not effective. In this sense, to our knowledge, there are no other multi-omic-oriented feature selection methods that satisfy the following requirements:

- Not specific to the HTS architecture
- Attributes \( P \) values to individual features to evaluate statistical significance
- Can handle more than or equal to three kinds of omics data simultaneously
- Applicable to severe large \( p \) small \( n \) (typically, \( p/n \sim 10^2 \) or more) problems

For example, for the large \( p \) small \( n \) problem, although Subramanian et al. [2] summarized existing machine learning methods for multi-omics analysis, typically they are applied to studies including up to \( 10^2 \) samples; thus, they cannot be regarded as a severe large \( p \) small \( n \) problem. As a result, the performance of the KTD-based unsupervised FE was not directly compared to other existing methods that satisfy all of the above conditions.

HBV vaccination data were selected to demonstrate the superior power of advanced KTD-based unsupervised FE because of the difficulty of the problem with this dataset. Since vaccination must be given to healthy people, side effects must be minimized [29]; in fact, since vaccination is essentially an infection with a weaker pathogen, its effect is inevitably weak. As expected, a very limited number of features (genes and proteins) were selected using the advanced KTD-based unsupervised FE proposed in this article, whereas conventional linear regression analysis did not attribute significant \( P \) values to any features. This suggests that the proposed advanced KTD-based unsupervised FE method has superior ability to select features when applied to even particularly difficult multi-omics datasets.

One might wonder why we did not employ more advanced feature selection methods other than lasso or rf. Generally, other methods are not fitted to the present situation, i.e., the large \( p \) small \( n \) problem. Since it is impossible to demonstrate the difficulty of using all the other methods, we consider two methods, class-specific mutual information variation for feature selection [30] and multilabel feature selection with constrained latent structure shared term [31] in order to demonstrate why other advanced methods are not fitted to the large \( p \) small \( n \) problem. When \( k \) features were aimed to be selected, although the complexity of class-specific mutual information variation for feature selection was supposed to be \( kMN \), it excluded the computational time needed for the computation of mutual information among \( N \) features, which is as large as \( N^2 \). It is not fitted to the large \( p \) small \( n \) problem associated with a large number of features, \( N \). For example, in the synthetic example, we tried to compute mutual information among \( N = 100 \) features for only one ensemble; it took 100 s. Since we employed \( N = 1000 \), it would take \( 100 \times (1000/100)^2 = 10^4 \) s for only one ensemble. We employed 100 ensembles; thus, in total, the required computational time for 100 ensembles would be as long as \( 10^4 \times 100 = \) one million s, which is unrealistic, since other methods, such as linear regression, lasso and rf, require less than 10 min (= 600 s) for computation with 100 ensembles. This means that class-specific mutual information variation for feature selection is not reasonable to be applied to the present synthetic example. As for multilabel feature selection with constrained latent structure shared term, it is not fitted to the large \( p \) small \( n \) problem as well, since it can select at most \( M \) features. Multilabel feature selection with constrained latent structure shared term decomposes the matrix \( N \times M \) into a product of two small matrices, \( N \times k \) and \( k \times M \), when \( k \) features are selected. Nevertheless, in the large \( p \) small \( n \) problem, since \( M \ll N \), \( k < M < N \). Thus, it can select at most \( M \) features. On the other hand, multilabel feature selection with constrained latent structure shared term was applied to the case where \( N < M \) [31]. In our synthetic example, the number of features to be selected, \( 2N_{20} \), is larger than \( M \). Thus, multilabel feature selection with constrained latent structure shared term cannot be used for the present synthetic example. Although
these are only two examples, most of the popular feature selection methods are not suitable for the large $p$ small $n$ problem, as shown for these two methods.

**Conclusion**

In this paper, an advanced KTD-based unsupervised FE method was introduced, which was modified to be applied to feature selection in multi-omics data analysis that is often very difficult, mainly based on the large $p$ small $n$ problem. The proposed method was successfully applied to a synthetic dataset, as well as to two real datasets, and attributed significant $P$ values to features with reduced CPU time and memory, even when applied to integrated analysis of more than two multi-omics datasets. Although the modification from the previously proposed KTD-based unsupervised FE was not significant, this slight modification was successful when applied to feature selection of multi-omics data analysis, which often poses a challenge in the case of a large $p$ small $n$ problem.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12920-022-01181-4.

- **Additional file 1.** Tables S1 to S4. Supplementary Tables.
- **Additional file 2:** Data S1. List of selected features and enrichment analysis for methylation.
- **Additional file 3:** Data S2. List of selected features and enrichment analysis for gene expression.
- **Additional file 4:** Data S3. List of selected features and enrichment analysis for proteome of WBC.
- **Additional file 5:** Data S4. List of selected features and enrichment analysis for proteome of Plasma.

**Acknowledgements**

Not applicable.

**Authors’ contributions**

YHT planned the research, performed analyses. YHT and TT have evaluated the results, discussions, outcomes and wrote and reviewed the manuscript. All authors read and approved the final manuscript.

**Funding**

This work was supported by Japan Society for the Promotion of Science, KAKENHI [Grant Nos. 19H05270, 20K12067, 20H04848] to YHT.

**Availability of data and materials**

TCGA data set can be downloaded from http://firebrowse.org/ (No accession number was assigned). GEO data set can be downloaded using GEO ID: GSE16441, GSE155198, and GSE161020. ProteomeXchange data set can be downloaded using ID PXD0020474.

**Declaration**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12920-022-01181-4.

**Additional file 1.** Tables S1 to S4. Supplementary Tables.

**Additional file 2:** Data S1. List of selected features and enrichment analysis for methylation.

**Additional file 3:** Data S2. List of selected features and enrichment analysis for gene expression.

**Additional file 4:** Data S3. List of selected features and enrichment analysis for proteome of WBC.

**Additional file 5:** Data S4. List of selected features and enrichment analysis for proteome of Plasma.

**References**

1. Reel PS, et al. Using machine learning approaches for multi-omics data analysis: a review. Biotechnol Adv. 2021;49:107739.
2. Subramanian I, et al. Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights. 2020;14:1177932219899051.
3. Huynh PH, et al. Improvements in the large $p$, small $n$ classification issue. SN Comput Sci. 2020;1(4):1–19.
4. Taguchi YH. Unsupervised feature extraction applied to bioinformatics. Berlin: Springer; 2020.
5. Taguchi Y-H, Turki T. Mathematical formulation and application of kernel tensor decomposition based unsupervised feature extraction. Knowl-Based Syst. 2021;217:106834.
6. Roy SS, Taguchi Y-H. Identification of genes associated with altered gene expression and mRNA profiles during hypoxia using tensor decomposition based unsupervised feature extraction. Sci Rep. 2021;11(1):8909.
7. Taguchi YH. Tensor decomposition-based and principal-component-analysis-based unsupervised feature extraction applied to the gene expression and methylation profiles in the brains of social insects with multiple castes. BMC Bioinform. 2018. https://doi.org/10.1186/s12859-018-2068-7.
8. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing (2020). https://www.R-project.org/.
9. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc: Statist Series B (Methodol). 1996;58(1):267–88.
10. Liaw A, Wiener M. Classification and regression by random forest. R News. 2002;2(3):18–22.
11. Deutsch EW, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. Nucleic Acids Res. 2016;45(D1):1100–6.
12. Ng KL, Taguchi Y-H. Identification of mRNA signatures for kidney renal clear cell carcinoma using the tensor-decomposition method. Sci Rep. 2020;10(1):1–11.
13. Kuleshov MV, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):90–7.
14. Huang DW, et al. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2008;4(1):44–57.
15. Li X, et al. Integrated analysis of MicroRNA (miRNA) and mRNA profiles reveals reduced correlation between MicroRNA and target gene in cancer. Biomed Res Int. 2018;2018:1–15.
16. Hong L, et al. Role of hippo signaling in regulating immunity. Cell Mol Immunol. 2018;15(12):1003–9.
17. Khalili M, et al. Diabetes and prediabetes in patients with hepatitis b residing in North America. Hepatology. 2015;62(5):1364–74.
18. Song C, et al. Associations between hepatitis B virus infection and risk of all cancer types. JAMA Netw Open. 2019;2(6):e195718.
19. Pakpour J, et al. Viral hepatitis and Parkinson disease. Neurology. 2017;88(17):1630–3.
20. Wang Y-R, et al. Associations between hepatic functions and plasma amyloid-beta levels-implications for the capacity of liver in peripheral amyloid-beta clearance. Mol Neurobiol. 2016;54(3):2338–44.
21. Ji M, et al. Hepatitis B core VLP-based mis-disordered tau vaccine elicits strong immune response and alleviates cognitive deficits and neuropathology progression in tau.p301s mouse model of Alzheimer’s disease and frontotemporal dementia. Alzheimer’s Res Ther. 2018;10(1):1–15.
22. Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):10544–47.
23. Tusher VG, et al. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci. 2001;98(9):5116–21.
24. Li J, et al. Multi-kernel linear mixed model with adaptive lasso for prediction analysis on high-dimensional multi-omics data. Bioinformatics. 2019;36(6):1785–94.
25. Lun ATL, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. 2015;44(S):45–45.
26. Argelaguet R, et al. Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14(6):8124.
27. Singh A, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019;35(17):3055–62.
28. Seoane JA, et al. A pathway-based data integration framework for prediction of disease progression. Bioinformatics. 2013;30(6):838–45.
29. Jacobson RM, et al. Making vaccines more acceptable - methods to prevent and minimize pain and other common adverse events associated with vaccines. Vaccine. 2001;19(17):2418–27.
30. Gao W, Hu L, Zhang P. Class-specific mutual information variation for feature selection. Pattern Recog. 2018;79:328–39. https://doi.org/10.1016/j.patcog.2018.02.020.
31. Gao W, Li Y, Hu L. Multilabel feature selection with constrained latent structure shared term. IEEE Trans Neural Netw Learn Syst. 2021. https://doi.org/10.1109/TNNLS.2021.3105142.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.