Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Systems biology and multiomics

OmiHier: simultaneous learning of class label hierarchy in omics data multiclass classification

Jiemin Xie[1], Guanghao Wu[2], Keyi Li[3], Zhanyu Liang[1], Bozhen Ren[1], Xuemei Liu[4], Yunhui Xiong[1], Li C. Xia[1,*]

[1] Department of Statistics and Financial Mathematics, School of Mathematics, South China University of Technology, Guangzhou 510000, China.

[2] School of Software Engineering, South China University of Technology, Guangzhou 510000, China.

[3] School of Professional Studies, Columbia University, NY 10027, USA

[4] School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510000, China

*Correspondence: Li C. Xia, email: lcxia@scut.edu.cn.

Abstract

Motivation: Class hierarchies are common in biological systems, such as the class label organizations of differentiating cells or evolving tumor clones. Knowing the hierarchy can be very informative in guiding multi-class classification of omics data. However, most existing omics classifiers either simply ignore the hierarchy to multi-classify the samples directly, or rely on a predefined hierarchy that risks subjectively misrepresenting the underlying cascade, leading to less accurate results.

Results: To overcome this issue, we propose Omics Hierarchy Learning (**OmiHier**), a data-driven class hierarchy learning algorithm. OmiHier employs a bottom-up iterative approach, interlacing classification error minimization with successive label merging, thus enables automatic simultaneous learning of class hierarchy and sample labels. We evaluated OmiHier on simulated and real-world datasets, including multi-omics data originating from complex disease, microbiome, single-cell and spatial transcriptomics scenarios. Our benchmark demonstrated OmiHier's high performance in both sample classification accuracy and inferring the true biological cascade.

Availability: An open-source R implementation of OmiHier is available at: https://github.com/labxscut/OmiHier.

Contact: lcxia@scut.edu.cn

1 Introduction

In multi-omics data classification, co-inference of class hierarchy has emerged as an effective approach to study the underlying complex biological systems. However, existing direct one-step multi-class classifiers or classifiers using subjectively defined class hierarchy lack the proper mathematical modelling of the space of potential hierarchies, which lead to inaccurate classification results and minimal understanding of the underlying biosystem. Integrating automated label hierarchy learning with the classifier training process may enhance sample classification performance, leading to more accurate and interpretable results. This is particularly relevant when the data are abundant with feature-level noises such as dropouts, outliers, and batch bias. Moreover, learning label hierarchy can improve the interpretability of resultant classification model. A well-learned class hierarchy can provide valuable insights into the relationships among biological entities, aiding researchers in comprehending

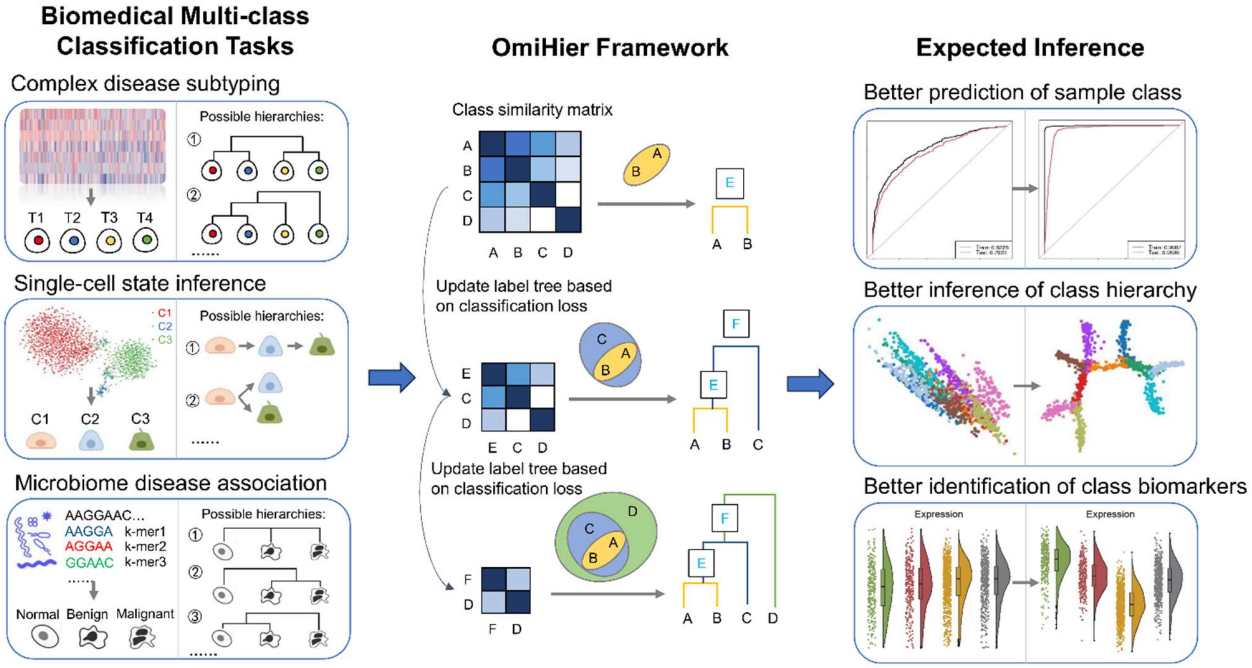

Fig 1. Design of this study, including multi-omics data multi-class classification tasks, the OmiHier iterative algorithm framework and the expected inference improvements.

the underlying biological processes and identifying potential biomarkers (Chen, et al., 2020; Knudsen, et al., 2018).

Despite the importance of also learning class hierarchy in multi-omics data classification tasks, current classifiers often overlook this aspect due to its associated complexities. For examples, Xie *et al.* demonstrated a significant improvement in the classification performance of large-scale cancer genomics data by employing a multi-step hierarchical classification scheme that aligns with the knowledge of the literature. In the context of classifying intrinsic subtypes of breast cancer, the classification accuracy of employing a correct hierarchical structure is significantly better than that of a simple one-step multi-class classification method (Xie, et al., 2023). And in molecular subtyping of gastric cancer, the classification accuracy based on a correct hierarchy is also better than that of a simple one-step multi-class classification method (Yang, et al., 2023 (accepted)). Consequently, there is a clear need for developing novel algorithms that simultaneously learn class label and hierarchy for omics data.

However, existing hierarchical learning approaches for the purpose have primarily been developed for image classification and natural language processing. For examples, P. Perona *et al.* (Griffin and Perona, 2008) and S. Bengio *et al.* (Bengio, et al., 2010) proposed the method of joint multiple One-vs-Rest classifiers. They constructed confusion matrices on the validation set and prioritizing the most difficult-to-separate categories from bottom up. Li *et al.* further constrained the hierarchical tree structure with a new loss function and generalized the hierarchical structure learning method to the case of non-mutually exclusive multi-label sets (Deng, et al., 2011). D. Casasent *et al.* proposed three methods for constructing hierarchical structures based on clustering: a method based on the balanced class partitioning (Casasent and Wang, 2005), another on the similarity of the original space (Wang and Casasent, 2006) and the third on the high-dimensional space (Wang and Casasent, 2007).

Leveraging on these developments and considering the unique properties of omics data, we propose Omics Hierarchy Learning (**OmiHier**), an automatic, data-driven hierarchical learning tool, accommodating the high-throughput, dimensionality, and sparsity of omics data (**Fig. 1**). OmiHier is based on an iterative framework, alternating between supervised updates of sample labels and unsupervised updates of label class clusters, all data driven. In the end, OmiHier automatically and simultaneously learns the most probable label hierarchy, as well as the hierarchy -associated optimal classifier collections for future predictions.

In the paper, we present the development and evaluation of OmiHier, with both simulated and a diverse range of real-world multi-omics datasets, from complex disease, microbiome, single-cell, and spatial transcriptomics scenarios. These results demonstrated OmiHier's capacity in multi-class classification tasks, not only outperforming traditional methods but also revealing meaningful label structures inherent to the data.

2 Methods

2.1 Omics Hierarchy Learning (OmiHier) algorithm

The OmiHier algorithm employs an iteratively bottom-up strategy. Starting from the elemental class level, the algorithm progressively aggregates and merges classes and moves to higher levels until all classes are merged. To minimize the overall hierarchical classification tree error, OmiHier performs three steps in the i-th iteration:

- Cross-validate the One-vs-Rest classifiers and then obtain the class label error confusion matrix using test set:

$$C_{a,b}^{(i)} = |\{(x_j, y_j) \in Y : y_j = a;\ argmax_{k\in\{1,\dots,K\}}\ f_k{}^{(i)}(x_j) = b\}| \quad (1)$$

  where $f_k{}^{(i)}$ is the cross-validated One-vs-Rest base classifier for class k, Y is the test set, K is the number of total classes.

- Construct an active class (or merged class) similarity matrix based on $C^{(i)}$:

$$A_j{}^{(i)} = \frac{C_{jk}{}^{(i)}}{\max\left(C_j{}^{(i)}\right)}; \qquad A^{(i)} = \frac{1}{2}\left(A^{(i)} + A^{(i)^T}\right) \quad (2)$$

- Merge the active classes that are most similar, to let them separate at lower levels, then move up one level and repeat the procedure until all classes have been merged to one (see **Algorithm 1**).

Algorithm 1: OmiHier bottom-up classification tree learning algorithm

---

Input: Affinity matrix $A^{(0)}$, the number of classes K, $T(N, E) = T(\{\{k\}: k = 1,\dots, K\}, \emptyset)$

Initialize: Node class label table: $B = (\{1\}, \dots, \{K\})$, active node (class) sets: $n_K = \{1,\dots, K\}$, $i = 0$

When $i < K$, **repeat**: // A total of $K - 1$ merges required

// Combine the most similar q, p nodes (classes)

$\{q, p\} = argmax_{\{a,b\}} A^{(i)}_{a,b}$, $a, b \in n_{i+K}$

// Add a new node (for the mega merged class)

$N = N \cup \{i + K + 1\}$

// Deactivate the merged nodes and activate the new node

$n_{i+K+1} = (n_{i+K} \backslash \{q, p\}) \cup \{i + K + 1\}$

// Add the new node's classes to the label set table

$B = (B, B[q] \cup B[p])$

// Add edges between the new node and its child nodes

$E = E \cup (p, i + K + 1) \cup (q, i + K + 1)$

Train $f^{(i+1)}$ to get the new confusion matrix $C^{(i+1)}$ and the new affinity matrix $A^{(i+1)}$ by Eqs. (1) & (2)

// Move up one level for the next merge

$i = i + 1$

Output: Hierarchical structure T, and the set of class or mega classes B corresponding to the nodes of T

2.2 Base classifiers

In the study, we used several base classifiers to perform multi-class classification tasks on simulated and real-world multi-omics datasets. For each task, we performed a 10-fold cross validation on the training data to learn the optimal parameterization. Subsequently, we constructed the best One-vs-Rest classifiers by applying them to the test data.

We applied a Lasso-logistic regression model to the BRCA and GC datasets (see **Table 1**). This model consists of Least Absolute Shrinkage and Selection Operator (Lasso) regression with the logistic regression method. Suppose the response variable has K levels $G = \{1, 2, \dots, K\}$, and each sample x_i in our study has m features, i.e., $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}$. Let Y be the $N \times K$ indicator response matrix, with elements $y_{il} = I(g_i = l)$. Then the Lasso elastic net penalized negative log-likelihood function for minimization can be expressed as:

$$l(\{\beta_{0k}, \beta_k\}_1^K) = -\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} y_{il}(\beta_{0k} + x_i^T\beta_k) - \log\left(\sum_{l=1}^{K} e^{\beta_{0l} + x_i^T\beta_l}\right)\right)\right] + \lambda\sum_{j=1}^{p}|\beta_j| \quad (3)$$

where β is a $p \times K$ matrix of coefficients, β_k refers to the kth column (for outcome category k), and β_j the jth row (vector of K coefficients for variable j), λ is the penalty parameter.

We applied the LightGBM (Ke, et al., 2017) base classifiers to the Faecal and HCC datasets. LightGBM is a robust gradient boosting framework. For the simulated, NCI-scRNA and Lymphoid datasets, we applied the Support Vector Machine (SVM) base classifiers (see section 2.5 for details).

2.3 Evaluation metrics

We evaluate classification performance using the Area Under the Receiver Operating Characteristic Curve (AUC), accuracy, F1 score, precision and recall. To evaluate the correctness of inferred class hierarchy, we use Robinson-Foulds (**RF**) distance - an error measuring the difference of inner branches between the inferred and the truth tree. The RF distance between two trees T_1 and T_2 with n tips is defined as:

$$d(T_1, T_2) = i(T_1) + i(T_2) - 2v_s(T_1, T_2) \quad (4)$$

where $i(T_1)$ denotes the number of internal edges and $v_s(T_1, T_2)$ denotes the number of internal splits shared by the two trees. The normalized RF distance is derived by dividing $d(T_1, T_2)$ by the maximal possible distance $i(T_1) + i(T_2)$. If all samples are correctly classified, the RF distance between the inferred and the true hierarchy is 0. The *Ape* package in R was used to compute the RF distance (Robinson and Foulds, 1981).

2.4 Simulation

As an initial test of OmiHier, we employed the following approach to generate the simulation data. First, we assumed a hierarchical structure represented by K multi-variate normal classes. The classes are balanced with a total number of samples S. The number of features is f. For each class $j \in \{1,2, \dots, K\}$, we first sample the mean of its feature data ($1 \times f$ vector), following an independent and identically distributed (i.i.d.) uniform distribution: $m_{jl} \sim \mathrm{U}(1/2\mu, 3/2\mu)$, where $l \in \{1,2,\dots, f\}$ denotes the feature index and $\mu \in \mathbf{R}$ is a fixed real number, so that the largest difference of feature means between two classes is bounded by μ. We denote the standard deviation shared by all classes as σ, where a larger σ means greater dispersion of the data. We let $\lambda = \sigma/\mu$, which is the mean coefficient of variation of the data. We then sample the feature data from multivariate normal distributions $d_{jl} \sim \mathrm{MVN}(m_j, \sigma^2 I)$, where m_j denotes the mean vector corresponding to class j. For simplicity, we denote the simulation's hyperparameter set as (K, S, f, μ, λ).

In the simulation, we mainly focused on evaluating the effect of the mean coefficient of variation λ (data noisiness) and the number of features f on OmiHier's performance. To assess the effect of data variability, we set $(K = \{3,4,6,10\},\ S = 10\text{^}3,\ f = 10\text{^}3,\ \mu = 1,\ \lambda = \{1/100, 1/10, 1/6, 1/4, 1/3, 1, 3\})$. We generated 28 simulated datasets representing a range of hyperparameter combinations.

2.5 Real-world data collection

We prepared six real-world multi-omics datasets for benchmark. Breast and gastric cancer data were downloaded from The Cancer Genome Atlas (TCGA) (Bonneville, et al., 2017) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (Curtis, et al., 2012), including mutation, copy number aberration and methylation data, and through both the cBioPortal (Cerami, et al., 2012) and the TCGA data portal. Colon cancer microbiome data was downloaded from the ENA database (PRJEB7774) (Feng, et al., 2015). Gastric cancer cell line NCI-N87 scRNA-seq data was downloaded from Gene Expression Omnibus (GSE142750) and National Institute of Health's SRA (PRJNA498809) (Andor, et al., 2020). Lymphoid cell scRNA-seq data was downloaded from human Ensemble Cell Atlas (hECA) system (Chen, et al., 2022). Liver cancer spatial transcriptome data was downloaded from the HCCDB database (Integrative Molecular Database of

Hepatocellular Carcinoma) (Jiang, et al., 2023). These datasets all come with sample- or cell-level multi-omics profiles and known biological labels, such as disease subtype and status, cell lineage and cell clonality.

3 Results

3.1 OmiHier achieved high performance on sample classification and hierarchy inference with simulated data

We conducted a series of experiments on simulated datasets. We used the mean coefficient of variation λ, i.e., the overall variability of data and the number of total classes K as surrogates for task difficulty. We found that OmiHier maintained high accuracy (AUC=1, see **Fig. 2a**) and the inferred hierarchy was always accurate (normalized RF distance=0, see **Fig. 2b**) with the increase of λ up to 1 (meaning feature-wise variance is the same scale as mean) for classifying up to $K=10$ classes. It was only when λ exceeded 1 that classification accuracy showed a slight drop. This suggested that OmiHier reliably inferred correct hierarchies and was robust to data variability. Simultaneously, it maintained highly accurate in classification in facing considerable data variability.

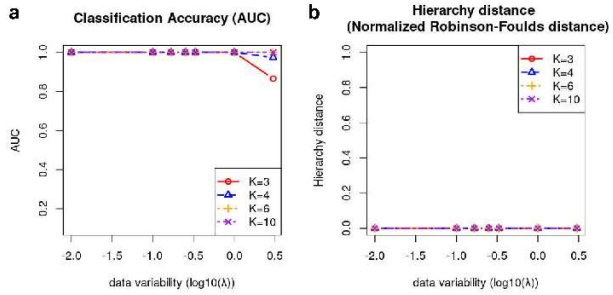

Fig 2. Classification and hierarchy inference performance of OmiHier on the simulated data. Classification accuracy (AUC score) and hierarchy correctness (normalized RF distance) were shown by task difficulties indicated by data variation ($\log10(\lambda)$) and class number (K).

3.2 OmiHier achieved high performance on sample classification and hierarchy inference with real-world data

We evaluated OmiHier on diverse real-world datasets, including complex disease subtyping, microbiome, single-cell and spatial transcriptomics data. We provided detailed information about each dataset, including the data type, the dataset size, and the subtype distribution in **Table 1**. As shown in **Table 2**, OmiHier consistently outperformed direct non-hierarchical classifiers, demonstrating its effectiveness in inferring meaningful class hierarchies and improving classification accuracy.

Specifically, for the combined breast cancer genomic and epigenomic (BRCA) subtyping problem, OmiHier achieved: AUC score of 0.980, accuracy of 0.883, F1 score of 0.893, precision of 0.883, and recall of 0.917, respectively. Notably, these results represented significant improvements over one-step direct multi-class classification, a strategy used by PAM50 (Parker, et al., 2009) – the current gold standard of BRCA subtyping, with an increase of 0.045 in AUC score, 0.17 in accuracy, 0.17 in F1 score, 0.17 in precision, and 0.16 in recall. For the combined gastric cancer genomic and epigenomic (GC) dataset, OmiHier achieved the following metrics: AUC score of 0.998, precision of 0.955, F1 score of 0.957, precision of 0.955, and recall of 0.963, all of which are also higher than the one-step direct multi-class classification approach.

Table 1. Real-world multi-omics datasets used in benchmark.

Dataset	Data type	Sample number	Feature number	Subtypes / Classes
BRCA	Breast cancer genomic and epigenomic data	2065	50831	Basal-like: 311 Her2-enriched: 249 Luminal A: 978 Luminal B: 527
GC	Gastric cancer genomic and epigenomic data	757	53498	CIN: 297 GS: 155 MSI: 159 EBV: 146
Faecal	Faecal microbiomes (in K-tuple/30-mer feature representation)	156	150	Normal: 63 Benign: 45 Malignant: 48
HCC	Liver cancer (HCC) spatial transcriptome data	8859	2000	Immune: 43 Normal: 3560 Stromal:1793 Tumor: 3463
NCI-scRNA	Single-cell RNA-seq in gastric cancer line NCI-N87	2168	1207	R1: 1337 R2: 128 R3: 703
Lymphoid	Single-cell RNA-seq of lymphoid cells	12067	20	CD4 T cell: 1825 NK cell: 4375 NK T cell: 3748 CD8 T cell: 1588 B cell: 531

Table 2. Benchmark results of OmiHier on real-world datasets.

Dataset	Classifier	Hierarchy	AUC	Accuracy	F1	Precision	Recall
BRCA	Lasso-Logistic	OmiHier	**0.980**	**0.883**	**0.893**	**0.883**	**0.917**
		direct	0.935	0.714	0.725	0.714	0.752
GC	Lasso-Logistic	OmiHier	**0.998**	**0.955**	**0.957**	**0.955**	**0.963**
		direct	0.997	0.950	0.952	0.950	0.961
Faecal	LightGBM	OmiHier (direct)	**0.857**	**0.677**	**0.667**	**0.677**	**0.670**
HCC	LightGBM	OmiHier	**0.978**	0.960	0.958	0.960	0.960
		direct	0.952	**0.964**	**0.963**	**0.964**	**0.964**
NCI-scRNA	SVM	OmiHier	**0.999**	0.979	**0.98**	**0.980**	**0.980**
		direct	0.998	0.982	0.971	0.972	0.973
Lymphoid	SVM	OmiHier	**0.982**	**0.887**	**0.886**	**0.887**	**0.886**
		direct	0.978	0.882	0.882	0.882	0.882

Note: in Hierarchy: direct for non-hierarchical multi-classifier. OmiHier denotes OmiHier learned hierarchical classifier, OmiHier (direct) means OmiHier inferred a non-hierarchical multi-classifier, indicating that OmiHier inferred is consistent with direct hierarchy.

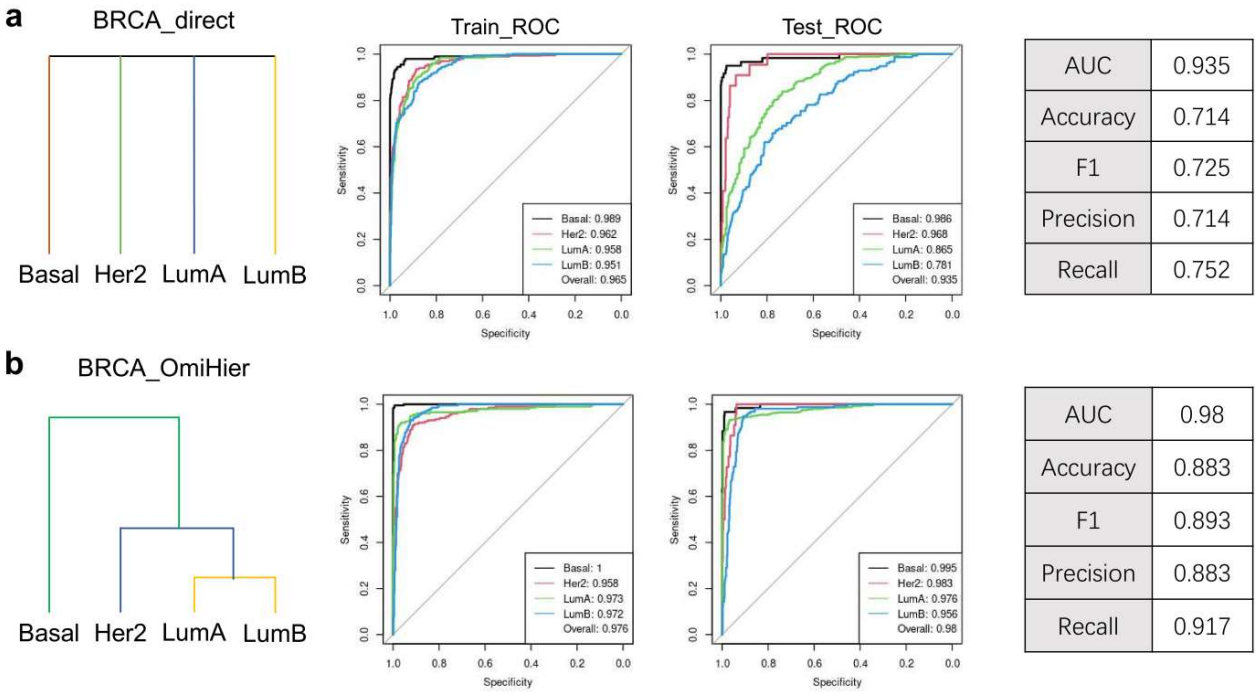

Fig 3. OmiHier breast cancer subtyping. The classification performance of **a.** best trained non-hierarchical multi-class classification model (BRCA_direct); **b.** OmiHier inferred hierarchical breast cancer subtype classifier (BRCA_OmiHier).

For the faecal microbiome (Faecal) dataset, the hierarchical classifier learned by OmiHier is consistent with the one-step multi-class classification and they both exhibit identical performance. For the liver cancer spatial transcriptome (HCC) dataset, OmiHier achieved: AUC score of 0.978, precision of 0.960, F1 score of 0.958, precision of 0.960, and recall of 0.960, respectively, and the AUC score was 0.026 higher than the non-hierarchical strategy, with other metrics being close (-0.005 at most).

For the single-cell RNA-seq of gastric cancer line NCI-N87 (NCI-scRNA) dataset, OmiHier achieved: AUC score of 0.999, precision of 0.979, F1 score of 0.980, precision of 0.980, and recall of 0.980, respectively, with four metrics higher than those of non-hierarchical multi-class classification, with the accuracy only 0.003 lower. For the single-cell RNA-seq of lymphoid cells (Lymphoid) dataset, OmiHier achieved: AUC score of 0.982, precision of 0.887, F1 score of 0.886, precision of 0.887, and recall of 0.886, respectively, all of which are higher than the one-step non-hierarchical classification strategy.

3.3 OmiHier breast cancer intrinsic subtype classifier

Breast cancer is a malignant, complex, and highly heterogeneous tumor that overwhelmingly impacts women. However, breast cancer subtyping remains a difficult scientific and clinical challenge (Horr and Buechler, 2021; Schettini, et al., 2022; Wolf, et al., 2022). The prevalent subtyping system discriminates among four primary intrinsic subtypes: Basal-like (Basal), Her2-enriched (Her2), Luminal A (LumA) and Luminal B (LumB). The system was established in the Prediction Analysis of Microarray of 189 patients based on the expression of 50 signature genes (i.e., PAM50 subtyping) (Parker, et al., 2009).

Clinically, intrinsic subtypes were determined by a combination of surrogate Immunohistochemistry tests (i.e., IHC surrogate) specifically targeting estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor 2 (Her2), and Ki67 proteins (Goldhirsch, et al., 2013; Goldhirsch, et al., 2011). Given their assessment difference, it is not surprising that PAM50 subtypes and IHC surrogates demonstrated significant inconsistencies and lack of interchangeability, as proven by many large-scale studies (Bastien, et al., 2012; Kim, et al., 2019). The situation, which could lead to mistreatment and misdiagnosis, call for a new identification methodology.

We applied OmiHier to investigate the intrinsic subtypes and their associated hierarchy with integrated genomic and epigenomic data from 2065 cases of breast cancer. The results of non-hierarchical classification model (BRCA_direct) were presented in **Fig. 3a**, while the results of OmiHier (BRCA_OmiHier) were presented in **Fig. 3b**.

We found the OmiHier learned BRCA_OmiHier hierarchical classifier to be most effective for identifying each subtype. In terms of the test dataset, the AUC scores are 0.995, 0.983, 0.976 and 0.956 for Basal, Her2, LumA and LumB subtypes, respectively, while the respective metrics for BRCA_direct classifier are 0.986, 0.968, 0.865 and 0.781, all considerably lower.

Historically, breast cancers have been broadly classified based on their gene expression profiles into Luminal- or Basal-type tumours. Only recently that the Luminal type was further divided into two sub-groups (Keller, et al., 2012). According to previous studies, the gene expression of LumA and LumB subtypes is similar (Jones, et al., 2004). Furthermore, during the clinical decision-making process, patients were first tested for Basal and Her2 biomarkers, followed by LumA and LumB

sub-markers for subtype classification (Prat, et al., 2015; Prat, et al., 2016).

The optimal subtype hierarchy learned by OmiHier (**Fig. 3b**) indeed showed that Luminal-like subtypes had greater difference from Basal and Her2 subtypes as they compared to each other. Therefore, we concluded that the BRCA_OmiHier hierarchy is a realistic classification tree model for identifying intrinsic subtypes also with higher accuracy. The non-hierarchical approach took by PAM50 based classifier or predefined hierarchy took by IHC, may explain their inaccuracy and inconsistency in clinical applications.

4 Conclusions

We developed the OmiHier algorithm and software tool that allows for simultaneous sample classification and inference of class hierarchy using multi-omics data. By applying OmiHier to large-scale multi-omics datasets, including cancer subtyping, microbiome, and single-cell lineage data, we demonstrated its high performance in classification accuracy and also in inferring the true underlying biological hierarchy. This leads to better identification of disease or cell subtypes and associated biomarkers. We hope that OmiHier finds broad application in studies of complex molecular hierarchies that drive biological systems.

Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation (2022A1515-011426 to LCX) and National Natural Science Foundation of China (61873027 to LCX).

Conflict of Interest: none declared.

References

Andor, N., *et al.* Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. *NAR Genomics and Bioinformatics* 2020;2(2):lqaa016.

Bastien, R.R.L., *et al.* PAM50 Breast Cancer Subtyping by RT-qPCR and Concordance with Standard Clinical Molecular Markers. *BMC Medical Genomics* 2012;5.

Bengio, S., Weston, J. and Grangier, D. Label embedding trees for large multi-class tasks. *Advances in Neural Information Processing Systems* 2010;23.

Bonneville, R., *et al.* Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol* 2017;1.

Casasent, D. and Wang, Y.-C. Automatic target recognition using new support vector machine. In, *Proceedings of the IEEE International Joint Conference on Neural Networks.*: IEEE; 2005. p. 84-89.

Cerami, E., *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2(5):401-404.

Chen, M., *et al.* Towards understanding hierarchical learning: Benefits of neural representations. *Advances in Neural Information Processing Systems* 2020;33:22134-22145.

Chen, S., *et al.* hECA: The cell-centric assembly of a cell atlas. *Iscience* 2022;25(5).

Curtis, C., *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486(7403):346-352.

Deng, J., *et al.* Fast and balanced: Efficient label tree learning for large scale object recognition. *Advances in Neural Information Processing Systems* 2011;24.

Feng, Q., *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nature communications* 2015;6(1):6528.

Goldhirsch, A., *et al.* Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013;24(9):2206-2223.

Goldhirsch, A., *et al.* Strategies for subtypes-dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011;22(8):1736-1747.

Griffin, G. and Perona, P. Learning and using taxonomies for fast visual categorization. In, *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2008. p. 1-8.

Horr, C. and Buechler, S.A. Breast Cancer Consensus Subtypes: A system for subtyping breast cancer tumors based on gene expression. *Npj Breast Cancer* 2021;7(1).

Jiang, Z., *et al.* HCCDB v2. 0: Decompose the Expression Variations by Single-cell RNA-seq and Spatial Transcriptomics in HCC. *bioRxiv* 2023:2023.2006. 2015.545045.

Jones, C., *et al.* Expression profiling of purified normal human luminal and myoepithelial breast cells: Identification of novel prognostic markers for breast cancer. *Cancer Res* 2004;64(9):3037-3045.

Ke, G., *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In, *Advances in Neural Information Processing Systems 30 (NIP 2017)*. 2017.

Keller, P.J., *et al.* Defining the cellular precursors to human breast cancer. *P Natl Acad Sci USA* 2012;109(8):2772-2777.

Kim, H.K., *et al.* Discordance of the PAM50 Intrinsic Subtypes Compared with Immunohistochemistry-Based Surrogate in Breast Cancer Patients: Potential Implication of Genomic Alterations of Discordance. *Cancer Res Treat* 2019;51(2):737-747.

Knudsen, T., Marchiori, D. and Warglien, M. Hierarchical decision-making produces persistent differences in learning performance. *Scientific Reports* 2018;8(1):1-12.

Parker, J.S., *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* 2009;27(8):1160-1167.

Prat, A., *et al.* Abstract P6-01-06: Feasibility of the PROSIGNA® multigene test in core biopsies and comparison to corresponding surgical breast cancer sections. *Cancer Res* 2015;75(9_Supplement):P6-01-06-P06-01-06.

Prat, A., *et al.* Prediction of Response to Neoadjuvant Chemotherapy Using Core Needle Biopsy Samples with the Prosigna AssayProsigna ROR Score Predicts Chemosensitivity. *Clinical Cancer Research* 2016;22(3):560-566.

Robinson, D.F. and Foulds, L.R. Comparison of phylogenetic trees. *Mathematical biosciences* 1981;53(1-2):131-147.

Schettini, F., *et al.* A perspective on the development and lack of interchangeability of the breast cancer intrinsic subtypes. *Npj Breast Cancer* 2022;8(1).

Wang, Y.-C. and Casasent, D. Hierarchical k-means clustering using new support vector machines for multi-class classification. In, *Proceedings of the IEEE International Joint Conference on Neural Network Proceedings*. IEEE; 2006. p. 3457-3464.

Wang, Y.-C. and Casasent, D. New weighted support vector k-means clustering for hierarchical multi-class classification. In, *Proceedings of the International Joint Conference on Neural Networks*. IEEE; 2007. p. 471-476.

Wolf, D.M., *et al.* Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell* 2022;40(6):609-+.

Xie, J.*, et al.* Building a genetic and epigenetic predictive model of breast cancer intrinsic subtypes using large-scale data and hierarchical structure learning. *BioRxiv* 2023:06.12.544702.

Yang, B.*, et al.* Hierarchical learning of gastric cancer molecular subtypes by integrating multi-modal DNA-level omics data and clinical stratification. *Quantitative Biology* 2023 (accepted).