Original Article

Associations of Normalization and Regularization with Machine Learning Overfitting in Cross-dataset Classification of Deaths Using Transcriptomic and Clinical Data: A Secondary Analysis of Publicly Available Databases

Fei Deng[1] and Lanjing Zhang[1,2,3,4*]

[1]Department of Chemical Biology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ, USA; [2]Department of Pathology, Princeton Medical Center, Plainsboro, NJ, USA; [3]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA; [4]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Background and objectives: Normalization can standardize and improve machine learning (ML) performance on omics data. However, it is unclear whether normalization is associated with overfitting (i.e., worse cross-dataset performance than intra-dataset performance). Therefore, we aimed to examine associations of normalization and regularization with overfitting of ML on omics data. ***Methods:*** Using three paired transcriptomic and clinical datasets (lung adenocarcinoma: the Cancer Genome Atlas (TCGA)/Oncology Singapore; melanoma: TCGA/Dana-Farber Cancer Institute; glioblastoma: TCGA/Clinical Proteomic Tumor Analysis Consortium), we applied ANOVA-based gene selection methods, six normalization methods, and six ML models to classify cancer patients' deaths. Balanced accuracy (BA) and area under the curve (AUC) in intra- and cross-dataset settings were compared using inferential analyses. ***Results:*** Normalization consistently improved intra-dataset performance (median BA/AUC changes: 0.035–0.214/0.115–0.279) on all data, particularly with Z_Raw, but decreased or slightly increased cross-dataset performance (median BA/AUC changes: −0.029–0.079/0.029–0.064). Least Absolute Shrinkage and Selection Operator (LASSO) model without normalization consistently outperformed most of the ML models in cross-dataset testing across cancer types. ML models on all and molecular-alone data showed similar best performances. ***Conclusions:*** Normalization increases ML's intra-dataset performance and overfitting in three paired cancer transcriptomic and clinical datasets. Regularized models such as LASSO appear to mitigate overfitting and achieve robust cross-dataset performance. Therefore, cross-dataset evaluation and regularized models are recommended to assess and reduce overfitting, while normalization should be used cautiously. Adding clinical data seems to have little impact on ML models' performance. However, future work on other diseases and datasets is warranted.

Introduction

Machine learning (ML) has become a cornerstone of bioinformatics, enabling predictive modeling for classification of diseases and patient outcomes using high-dimensional omics data.[1–4] It is particularly helpful in the era of massive production and application of high-throughput data.[5–7] However, the generalizability of ML models across datasets remains a critical challenge due to heterogeneity in experimental platforms, sample populations, and preprocessing techniques, reaching an F1 of 61% or an area under the curve (AUC) of the receiver-operating curve of 71% (dropped from 91% in intra-dataset testing) in cross-dataset testing.[8–11] ML models may indeed exhibit performance biases for sociodemographic groups.[12] Normalization is often assumed to enhance model performance.[7,13–17] However, its impact on cross-dataset performance is largely unknown, particularly for high-dimensional omics data where overfitting risks are high.[18–20]

A known cause of the ML generalizability problem is the possible over-reliance on intra-dataset cross-validation for model evaluation and selection.[21,22] While valuable in many cases, it suffers from selection bias and leads to overly optimistic estimates of a model's true performance.[8,19,20,23] Moreover, preprocessing strategies, such as data normalization and aggressive feature selection, can improve performance metrics within a single dataset,[24–26] but may unintentionally cause model overfitting. This intensive optimization can paradoxically harm the model's ability to generalize, a finding that has been noted in recent studies.[13,27] Finally,

feature selection methods (e.g., differentially expressed gene (DEG)) can improve intra-dataset performance but may exacerbate overfitting in cross-dataset validation.[19] The evaluation of ML performance also faces scrutiny, as intra-dataset metrics often fail to predict cross-dataset generalizability.[16,28] However, the association between preprocessing methods and ML's cross-dataset performance is unclear.

Regularization techniques, such as the Least Absolute Shrinkage and Selection Operator (LASSO),[29,30] have shown promise in reducing overfitting by penalizing model complexity, but their interaction with normalization remains poorly understood in classifying omics data. Recent studies suggest that simpler ML models may outperform complex methods in transcriptomics due to robustness to data variability.[27] However, it is largely unknown whether LASSO or other simple ML algorithms retain their performance in cross-dataset testing.

Therefore, we investigated the impact of normalization, regularization, and evaluation strategies on ML performance in classifying cancer deaths, focusing on cross-dataset performance. Using three pairs of transcriptomic and clinical datasets, we explored whether normalization can universally improve performance, assessed the impact of regularization, and evaluated the trade-offs of preprocessing and feature selection techniques. Our study may help develop robust ML pipelines with better generalizability in precision medicine and multi-omics applications.[31]

Materials and methods

Workflow and dataset selection

We searched for cancer transcriptomic datasets with clinical and death data in cBioPortal,[32] that also had at least one matched dataset with clinical and death data and could be used for independent cross-dataset testing. Three pairs of transcriptomic and clinical datasets in cancer were identified and used, including those of lung adenocarcinoma in the Cancer Genome Atlas (TCGA) and Oncology Singapore (OncoSG),[33,34] those of melanoma in TCGA and Dana-Farber Cancer Institute,[35,36] and those of glioblastoma in TCGA and the Clinical Proteomic Tumor Analysis Consortium.[36,37]

Specific experimental steps were described previously and repeated in all three pairs of cancer datasets (Fig. 1).[13] Briefly, 90% of randomly selected samples from the training dataset were used for training with five-fold cross-validation, while the remaining 10% served as an internal test set. Then the other dataset was used for cross-dataset testing, and vice versa. The entire process was repeated at least five times. Basic modeling factor values and key model hyperparameter settings were employed across all experimental steps of each process, including data cleaning, dataset partitioning, gene selection, normalization, classification model training, prediction, classification performance evaluation, and statistical analysis (Supplementary Table 1). Python version 3.11.9 64-bit was used for the code implementation.

The classification outcome/label was binary (living versus deceased) in all three pairs of datasets. Only the features shared by the training and testing datasets were used for model training and testing. After applying the sample inclusion and exclusion criteria (Fig. 2), all remaining samples with paired transcriptomic and clinical data were carried forward to the downstream workflow. Transcriptomic data are in RNA-seq FPKM format and are further normalized using Z-transformation. Some datasets, such as the TCGA and OncoSG lung adenocarcinoma datasets, are class-imbalanced. For binary classification, sample numbers with living and deceased are 212:74 in TCGA (total 286), and 125:42 in On-

coSG (total 167). The same 4:1 split (i.e., 80% for training and 20% for intra-dataset testing) was applied to the melanoma and glioblastoma datasets.

Data cleansing

To enable analyses for two datasets, we cleaned the samples by retaining only those with matching labels, keeping shared gene features, and filling missing values feature-wise across molecular data with training-set medians. After this preprocessing, the dataset of lung adenocarcinoma included 16,196 gene features and four clinical features: age, gender, tumor stage, and tumor mutational burden. These features were chosen because they are shared between the two datasets. Some features are numerical, while others are categorical, requiring tailored processing methods.

Gene selection

As in nearly all transcriptomic studies, the number of samples is significantly smaller than the number of features (e.g., 16,196 genes in lung adenocarcinoma datasets), leading to potential multicollinearity and an increased risk of overfitting. Therefore, feature selection was performed with ANOVA, as shown before,[3,4,13,38–40] while the F-value, which measures the ratio of these variances, was used to test the null hypothesis that all group means are equal.[13] By setting different thresholds of P-values, gene sets can be defined accordingly. For example, genes with P-values below a selected threshold are designated as DEGs for classification, while those above a chosen threshold are designated as non-differentially expressed genes (NDEGs) for normalization. Specifically, gene selection was performed using the training set only, and the selected feature sets (DEGs and NDEGs) were then fixed and directly applied to the internal and cross-dataset testing sets.

Normalization

To evaluate model generalizability on independent external cohorts and avoid information leakage across cohorts, we focused on a set of classical normalization strategies that can be applied without joint modeling across cohorts. Since the transcriptomic data used here were already Z-transformed, we first examined the effect of classification on both the original dataset (Z_Original) and the gene-filtered dataset (Z_Raw data). We then evaluated binarization (Z_Binary) and four other reference gene-based normalization methods applied to Z_Raw data: Non-Parametric Normalization (Z_NPN), Quantile Normalization (Z_QN), Quantile Normalization with Z-Score (Z_QNZ), and Normalization using Internal Control Genes (Z_NICG), as described before (Supplementary Table 2).[13,15,41–43] Each normalization method was applied independently to training, internal test, and external test datasets.

ML models

We trained six commonly used ML classifiers on different training sets using specific hyperparameter tuning settings (Supplementary Table 1), including multilayer perceptron,[44,45] extreme gradient boosting (XGB),[46,47] logistic regression,[48] LASSO,[29] support vector machine (SVM),[49] and random forest.[50] Considering the imbalance in the dataset, class weights were applied in the XGB and SVM models, referred to as XGB_W and SVM_W, respectively.

Classification performance evaluation

Due to the binary and unbalanced nature of the data in this study, balanced accuracy (BA) was the primary performance metric and AUC was the secondary.[21,22] We also used the median of the changes (delta) in model performance (versus
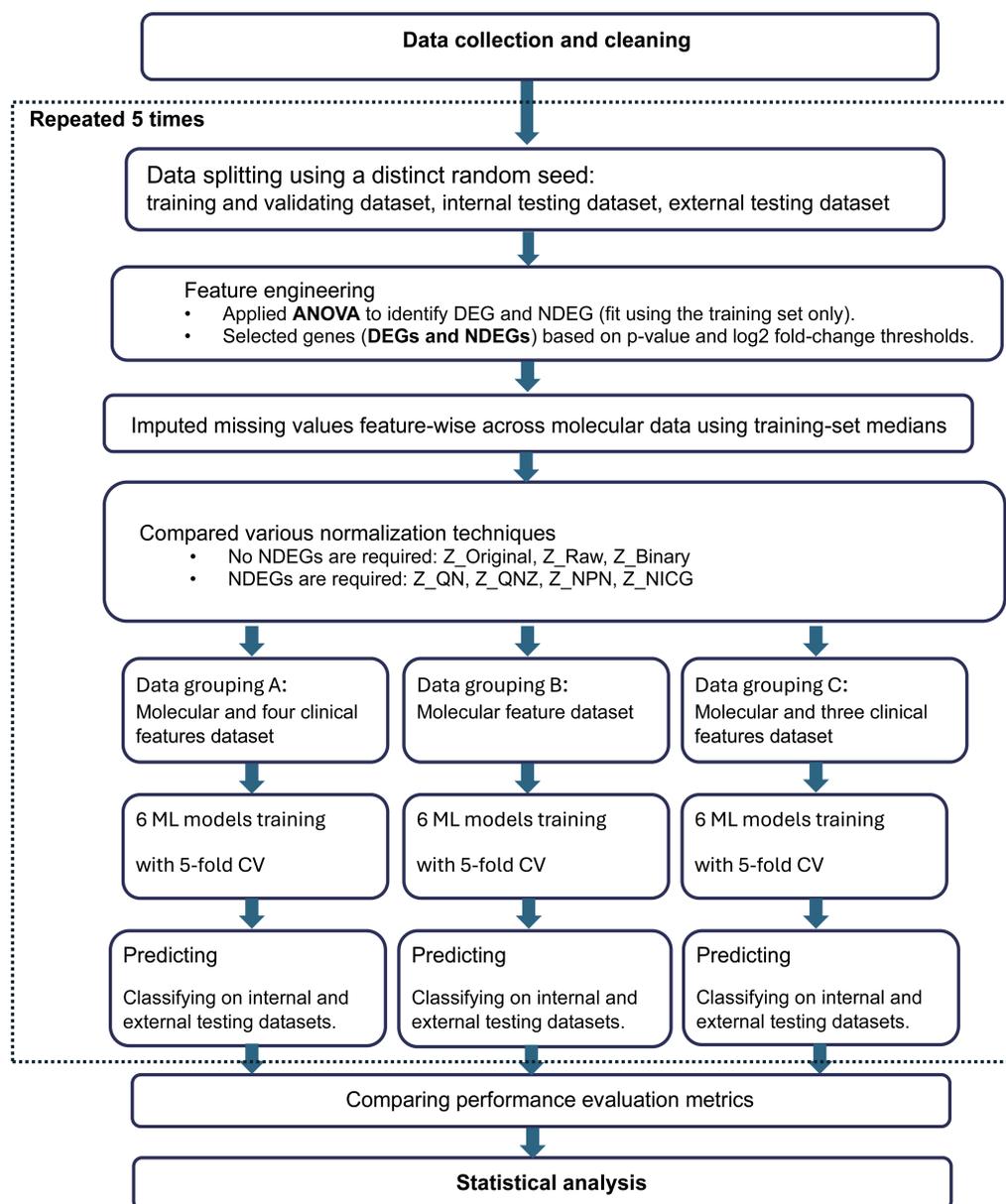
Fig. 1. The workflow that was repeated for each of the three cancer types, including lung adenocarcinoma, melanoma, and glioblastoma. CV, cross-validation; DEGs, differentially expressed genes; ML, machine learning; NDEGs, non-differentially expressed genes.

Z_Original) to evaluate the impact of normalization methods on the changes in model performance. A *P*-value less than 0.05 was considered statistically significant.

Statistical analysis

A layered statistical analysis framework was adopted for model performance. Following our previous work in Ref.13,28, we first constructed internal- and external-test "mean performance matrices" indexed by combinations of DEG and NDEG thresholds. The optimal value in the matrix was used as the representation for the corresponding model-normalization combination.

The first layer of analysis was based on the underlying repeated-run results corresponding to each representation (five repetitions for the internal test and 15 repetitions for the external test). In order to evaluate the benefit of incorporating clinical features during training, we applied Welch's t-test to compare model performance under each model-normalization combination with versus without clinical features.[51,52] In the second layer analysis, to assess whether feature selection and subsequent normalization improved model performance, we performed within-model paired comparisons of Z_Original and the other five normalization methods against the reference Z_Raw using Welch's t-test. The third layer analysis was only for lung adenocarcinoma datasets. To examine the impact of training-set choice on performance and cross-dataset generalization, Welch's t-test was also used to compare the optimal internal-test results (also the optimal external-test results) obtained when using TCGA versus OncoSG as the training set.

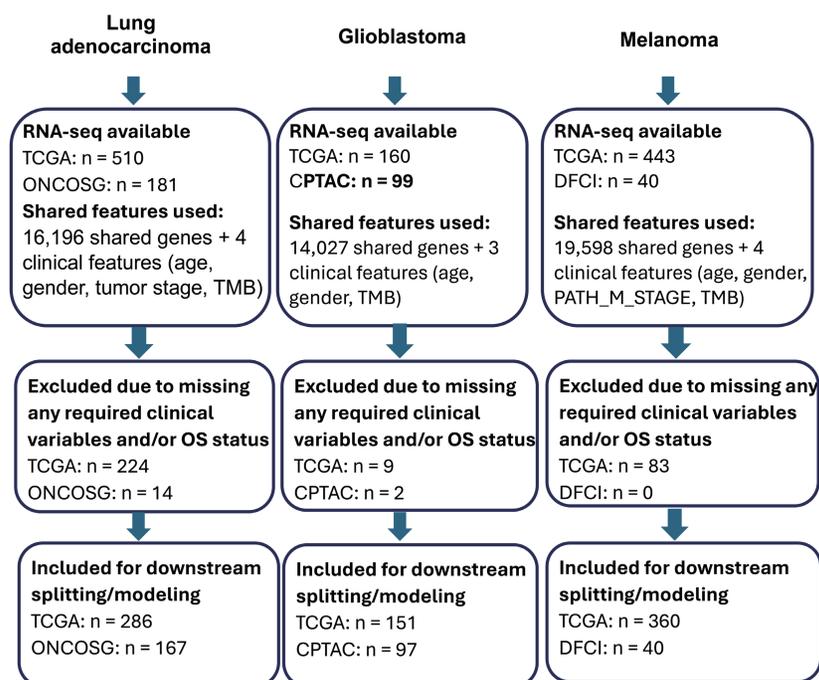**Fig. 2. Sample selection flow diagram.** CPTAC, the Clinical Proteomic Tumor Analysis Consortium; DFCI, Dana-Farber Cancer Institute; OncoSG, Oncogenomic-Singapore; OS, overall survival; PATH_M_STAGE, pathologic distant metastasis (M) stage; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden.

The fourth layer analysis was conducted on multiple "optimal model performance tables" generated under different training set choices and clinical features settings using the Wilcoxon signed-rank test.[53,54] Two paired tests were included: (1) row-wise comparison of Z_Original and the other normalization methods against Z_Raw; (2) column-wise comparison of the other models against LASSO. These analyses were used to evaluate the generalizability of normalization, feature selection, and model effects across different settings.

For each predefined comparison family, we controlled multiplicity by performing false discovery rate correction via the Benjamini-Hochberg procedure (q = 0.05). For layers 1–3, our primary goal was to compare mean performance across independent conditions. Because heteroscedasticity and unbalanced sample sizes might arise across repeated runs under different settings, we used Welch's t-test for two-group comparisons.[51,52] For layer 4, because comparisons involved greater differences in settings and distributional assumptions were harder to satisfy, we used the nonparametric paired Wilcoxon signed-rank test to compare paired differences.[53,54]

Results

Baseline characteristics of the datasets

The datasets all included transcriptomic and clinical data (Supplementary Tables 3–5). The outcome was binary living status. For lung adenocarcinoma, there were 212 alive and 74 deceased patients in the TCGA dataset (total 286) and 125 alive and 42 deceased in the OncoSG dataset (total 167) at the end of their follow-ups. For glioblastoma, there were 52 alive and 99 deceased patients in the TCGA dataset (total 151) and 35 alive and 62 deceased in the Clinical Proteomic Tumor Analysis Consortium dataset (total 97). For melano-

ma, there were 173 alive and 187 deceased patients in the TCGA dataset (total 360) and 13 alive and 27 deceased in the Dana-Farber Cancer Institute dataset (total 40).

Performances of ML models on lung adenocarcinoma data

We analyzed models' performances under various conditions, including multiple randomly selected sample combinations from the internal or external test sets. The best-performing models, when present, had statistically better BA and/or AUC than the average performance of all models (Supplementary Tables 6–21).

Models trained on the TCGA dataset and the OncoSG dataset exhibited different performances in external datasets. We then compared the best internal testing performances of models trained on the TCGA dataset with those trained on the OncoSG dataset under the three conditions mentioned above. When only transcriptomic data was used, the performance differences between the two datasets using the same method were statistically significant (Table 1). Moreover, the statistical significance of this difference was even more pronounced in cross-platform external testing. Models trained on the TCGA dataset showed significantly better predictive performance on the OncoSG dataset than the models trained on the OncoSG dataset when tested on the TCGA dataset. This discrepancy may stem from the fact that the OncoSG dataset primarily consists of samples from Asian populations.

For narrative convenience, we referred to the model based on genetic features and four clinical features as Data grouping A, the model using only genetic feature data as Data grouping B, and the one based on genetic features and three clinical features as Data grouping C.

We also compared the best performance of ML in internal testing and that in external testing obtained for data groupings A, B, and C (Supplementary Tables 22 and 23). Inter-

Table 1. Comparison of the best internal testing performance of models trained on the TCGA dataset (n = 510) versus those trained on the OncoSGdataset (n = 181).

	All data			Molecular data alone			All data except tumor stage		
	TCGA as training set	OncoSG as training set	FDR-adjusted P-value	TCGA as training set	OncoSG as training set	FDR-adjusted P-value	TCGA as training set	OncoSG as training set	FDR-adjusted P-value
Intra-dataset testing									
Balanced accuracy	0.814 ± 0.010	0.935 ± 0.004	0.179	0.848 ± 0.001	0.977 ± 0.000*	0.180	0.853 ± 0.011	0.927 ± 0.003	0.480
AUC	0.888 ± 0.023	0.953 ± 0.002	0.233	0.925 ± 0.019	1.000 ± 0.000*	0.180	0.885 ± 0.008	0.912 ± 0.010	0.892
Accuracy	0.821 ± 0.006	0.977 ± 0.001	0.076	0.890 ± 0.001	0.965 ± 0.001	0.180	0.910 ± 0.005	0.941 ± 0.002	0.107
DEG, n (p threshold)	78 (0.2%)	534 (0.4%)		1,430 (5%)	2,382 (4%)		996 (2%)	2,070 (2%)	
NDEG, n (p threshold)	62 (99%)	230 (95%)		120 (98%)	65 (99%)		120 (98%)	65 (99%)	
Normalization method	Z-Raw	Z-Raw		Z-NPN	Z-NICG		Z-Raw	Z-NICG	
Classification model	SVM_W	MLP		MLP	SVM_W		LR	MLP	
Cross-dataset testing									
Balanced accuracy	0.645 ± 0.003	0.556 ± 0.000*	0.003	0.657 ± 0.001	0.571 ± 0.000*	<0.001	0.654 ± 0.001	0.569 ± 0.000*	<0.001
AUC	0.654 ± 0.002	0.579 ± 0.000*	<0.001	0.687 ± 0.001	0.599 ± 0.000*	0.134	0.665 ± 0.002	0.595 ± 0.000*	0.001
Accuracy	0.645 ± 0.003	0.556 ± 0.000*	0.003	0.657 ± 0.001	0.571 ± 0.000*	<0.001	0.654 ± 0.001	0.569 ± 0.000*	<0.001
DEG, n (p threshold)	161 (0.6%)	617 (0.4%)		176 (0.7%)	2,382 (4%)		816 (1%)	2,960 (5%)	
NDEG, n (p threshold)	120 (98%)	1,729 (85%)		120 (98%)	230 (95%)		120 (98%)	562 (92%)	
Normalization method	Z-Binary	Z-Binary		Z-NICG	Z-QN		Z-Binary	Z-Binary	
Classification model	SVM_W	LR		LR	SVM_W		SVM_W	SVM_W	

All data are shown as mean ± standard deviation. All experiments were repeated 15 times. *The SD was less than 0.00001. AUC, area under the curve; DEGs, differentially expressed genes; FDR, False Discovery Rate; LR, Logistic Regression; MLP, Multilayer Perceptron; NDEGs, non-differentially expressed genes; SD, standard deviation; SVM_W, class weights were applied in the Support Vector Machine model; TCGA, The Cancer Genome Atlas; Z_Binary, binarization applied to Z_Raw data; Z_NICG, Normalization using Internal Control Genes (NICG) applied to Z_Raw data; Z_NPN, Non-Parametric Normalization (NPN) applied to Z_Raw data; Z_Original, Z-transformed RNA-seq data in FPKM format, including all gene features shared between the two cohorts; Z_QN, Quantile Normalization (QN) applied to Z_Raw data; Z_QNZ, Quantile Normalization with Z-Score (QNZ) applied to Z_Raw data; Z_Raw, Z_Original data restricted to the selected DEGs;

estingly, no models exhibited statistically significant differences, while the prediction performance of models trained on OncoSG data and applied to TCGA data showed significant differences under the three conditions.

Modelling with data in three cancer types

In intra-dataset testing across three cancer types (Table 2 and Supplementary Fig. 1), normalization methods consistently improved model performance compared to the reference Z_Original (no normalization after initial Z-score transformation). Improvements were substantial, as reflected in BA and AUC, which were as high as 0.814 and 0.889 in lung adenocarcinoma, 0.756 and 0.807 in melanoma, and 0.803 and 0.887 in glioblastoma, respectively. Across all cancer types, normalization markedly enhanced intra-dataset predictive performance for death classification, with Z_Raw often providing the greatest median improvement in glioblastoma and competitive gains in the other cancers. The performances of ML models using Z_Original and the other five normalization methods appeared overall better than those using Z_Raw (as the reference), as shown by Welch's t-test (Supplementary Tables 24–27).

In contrast to intra-dataset results, cross-dataset (external) testing revealed limited benefits from normalization and, in several cases, performance comparable to or worse than Z_Original (Table 3 and Supplementary Fig. 2).

A striking pattern emerged for the LASSO model. In cross-dataset testing of all three cancer types, LASSO achieved positive or minimally negative deltas across nearly all normalization methods, frequently outperforming the normalized versions of more complex models.

Overall, while normalization and associated gene selection markedly boosted intra-dataset performance, these preprocessing steps provided only marginal gains in cross-dataset testing and occasionally led to reduced performance. Differences in model performance were more pronounced in cross-dataset settings than within the same dataset, highlighting greater sensitivity to dataset heterogeneity in cross-dataset validation. Simpler, regularized approaches such as LASSO demonstrated consistent cross-dataset robustness, whereas more complex models showed variable and sometimes diminished generalizability after extensive normalization.

Modelling with molecular data alone in three cancer types

In intra-dataset testing using only molecular (transcriptomic) features across the three cancer types (Table 4 and Supplementary Fig. 3), normalization methods again substantially improved performance relative to Z_Original, with gains observed in both BA and AUC. Overall, normalization markedly enhanced intra-dataset death classification when using molecular data alone, with Z_Raw frequently delivering the strongest median gains, particularly in lung adenocarcinoma and glioblastoma.

Cross-dataset testing using only molecular features showed more limited and inconsistent benefits from normalization, similar to patterns observed with combined data, though overall performance levels were generally lower (Table 5 and Supplementary Fig. 4).

Across the three cancer types, normalization offered only marginal benefits in cross-dataset settings when relying solely on molecular data, and certain methods (e.g., Z_Binary) occasionally degraded performance. As with combined features, LASSO without additional normalization demonstrated remarkable consistency, achieving positive or near-neutral deltas in nearly all cross-dataset scenarios. In contrast, more complex models exhibited greater variability, underscoring

the robustness of regularized approaches for generalization across heterogeneous datasets.

Discussion

This large and comprehensive study using three pairs of cancer transcriptomic and clinical datasets reveals critical insights into ML performance in bioinformatics, particularly for cross-dataset generalization and preprocessing strategies. These results challenge conventional practices and may help develop robust, generalizable models for applications such as gene expression analyses.

First, we showed that the LASSO method without normalization consistently performed well in cross-dataset (external) testing of three pairs of transcriptomic and clinical datasets (i.e., three cancer types), challenging the necessity of normalization for cross-dataset tasks. This suggests that regularization inherent in LASSO effectively mitigates overfitting, simplifies preprocessing pipelines, and reduces computational costs.[29,55] Our finding also aligns with a study showing robust LASSO performance without extensive preprocessing.[27] However, LASSO may not be robust for some tasks, as shown in one study on spatial gene expression in the brain.[56]

Second, normalization and gene selection (DEG/NDEG) significantly improve intra-dataset performance but yield limited gains in cross-dataset testing, often leading to overfitting.[13,19,23–25] This underscores the need for cautious application of extensive preprocessing to avoid models that fail in external validation.[20] Interestingly, as shown by us and others, LASSO and other regularized methods can be used to reduce overfitting in microarray and single-cell RNA-seq data.[24,26]

Third, performance differences among ML models are smaller in intra-dataset testing than in cross-dataset settings, partly due to limited normalization and gene selection benefits in external contexts. This highlights the importance of selecting robust algorithms for heterogeneous datasets, such as random forest.[13,57–59] Indeed, recent works confirm that simpler, regularized models often maintain consistent performance across datasets, unlike complex models prone to overfitting.[60,61] This finding may help select models for applications requiring broader generalizability.[1]

Fourth, reliance on intra-dataset evaluation (e.g., cross-validation) may overestimate model generalizability, as shown by others and by us.[9,10] For example, the performance of ML models achieved with negative data generation cannot be transferred to cross-dataset testing, either.[9] We thus advocate shifting toward cross-dataset evaluation to prioritize models with consistent, acceptable performance, enhancing applicability in clinical settings like precision medicine,[8] while it is noteworthy that intra-dataset evaluation may match that of cross-dataset evaluation in some scenarios.[62] This paradigm shift addresses the gap between intra-dataset optimization and real-world robustness.[59] Others have also introduced benchmark datasets to robustly assess models' performance.[63] Further studies are needed to address this issue in more depth.

Finally, normalization's impact varies by ML model and is more pronounced with all data than with molecular data alone. This is particularly relevant for multi-omics integration, where data-specific preprocessing strategies are critical.[3,38,64] Recent studies on multi-omics data modeling support tailored normalization approaches to improve ML performance.[31,65,66] Therefore, we recommend data-specific ML workflows to enhance cross-dataset robustness.

Certain limitations of this study should be acknowledged.

Table 2. Intra-dataset testing of death classification on all data in three cancer types

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
BA													
Lung adenocarcinoma													
Z_Original	0.500(W)	Ref	0.500(W)	Ref	0.640(W)	Ref	0.500(W)	Ref	0.665**(W)	Ref	0.580*(W)	Ref	Ref
Z_Raw	0.570(B)	**0.070**	0.570(B)	**0.070**	0.740	**0.070**	0.570(B)	**0.070**	0.814(B)	**0.149**	0.698	0.118	**0.085**
Z_Binary	0.525	0.025	0.525	0.025	0.685	0.025	0.525	0.025	0.774	0.109	0.685	0.105	0.035
Z_NICG	0.563	0.063	0.563	0.063	0.792(B)	**0.152**	0.563	0.063	0.765	0.100	0.672	0.092	0.078
Z_NPN	0.525	0.025	0.525	0.025	0.770	0.130	0.525	0.025	0.790	0.125	0.696	0.116	0.071
Z_QN	0.538	0.038	0.538	0.038	0.755	0.115	0.538	0.038	0.782	0.117	0.687	0.107	0.073
Z_QNZ	0.550	0.050	0.550	0.050	0.783	0.143	0.550	0.050	0.757	0.092	0.709(B)	**0.129**	0.071
AUC													
Z_Original	0.607(W)	Ref	0.607(W)	Ref	0.754(W)	Ref	0.607(W)	Ref	0.645**(W)	Ref	0.656*(W)	Ref	Ref
Z_Raw	0.776	0.169	0.776	0.169	0.857(B)	0.169	0.776	0.169	0.888	0.243	0.806	0.150	0.169
Z_Binary	0.785	0.178	0.785	0.178	0.796	0.178	0.785	0.178	0.845	0.200	0.779	0.123	**0.178**
Z_NICG	0.742	0.135	0.742	0.135	0.838	0.135	0.742	0.135	0.889(B)	**0.244**	0.774	0.118	0.135
Z_NPN	0.754	0.147	0.754	0.147	0.852	0.147	0.754	0.147	0.871	0.226	0.783	0.127	0.147
Z_QN	0.756	0.149	0.756	0.149	0.842	0.149	0.756	0.149	0.838	0.193	0.788	0.132	0.149
Z_QNZ	0.787(B)	**0.180**	0.787(B)	**0.180**	0.836	**0.180**	0.787(B)	**0.180**	0.817	0.172	0.812(B)	**0.156**	0.176
BA													
Melanoma													
Z_Original	0.595(W)	Ref	0.573*(W)	Ref	0.588(W)	Ref	0.582(W)	Ref	0.575(W)	Ref	0.543(W)	Ref	Ref
Z_Raw	0.704	0.109	0.705	0.132	0.706	0.118	0.661	0.079	0.699	0.124	0.681	0.138	0.121
Z_Binary	0.699	0.104	0.680	0.107	0.665	0.077	0.678	0.096	0.665	0.09	0.687	0.144	0.100
Z_NICG	0.674	0.079	0.712	0.139	0.691	0.103	0.682	0.103	0.728(B)	**0.153**	0.708(B)	**0.165**	0.121
Z_NPN	0.690	0.095	0.715	0.142	0.714	0.126	0.689	0.107	0.725	0.15	0.674	0.131	0.129
Z_QN	0.713(B)	**0.118**	0.756(B)	**0.183**	0.719(B)	**0.131**	0.693	0.111	0.711	0.136	0.687	0.144	**0.134**
Z_QNZ	0.692	0.097	0.719	0.146	0.711	0.123	0.707(B)	**0.125**	0.706	0.131	0.665	0.122	0.124
AUC													
Z_Original	0.625(W)	Ref	0.605(W)	Ref	0.604(W)	Ref	0.635(W)	Ref	0.610(W)	Ref	0.578(W)	Ref	Ref
Z_Raw	0.767(B)	**0.142**	0.788	0.183	0.755	0.151	0.722	0.087	0.774	0.164	0.727	0.149	0.150
Z_Binary	0.738	0.113	0.752	0.147	0.735	0.131	0.731	0.096	0.705	0.095	0.746	0.168	0.122
Z_NICG	0.714	0.089	0.780	0.175	0.762	0.158	0.736	0.101	0.807(B)	**0.197**	0.770(B)	**0.192**	**0.167**
Z_NPN	0.739	0.114	0.778	0.173	0.794(B)	**0.19**	0.746	0.111	0.797	0.187	0.728	0.15	0.162

(continued)

Table 2. *(continued)*

Normaliza-tion method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_QN	0.766	0.141	0.794(B)	**0.189**	0.774	0.17	0.749	0.114	0.775	0.165	0.735	0.157	0.161
Z_QNZ	0.761	0.136	0.789	0.184	0.759	0.155	0.758(B)	**0.123**	0.778	0.168	0.737	0.159	0.157
Glioblastoma													
BA													
Z_Original	0.525***(W)	Ref	0.519***(W)	Ref	0.586**(W)	Ref	0.557(W)	Ref	0.519***(W)	Ref	0.552(W)	Ref	Ref
Z_Raw	0.761(B)	**0.236**	0.782	0.263	0.777(B)	**0.191**	0.604(B)	**0.047**	0.782	0.263	0.626	0.074	**0.214**
Z_Binary	0.650*	0.125	0.638**	0.119	0.655*	0.069	0.557(W)	0	0.578*	0.059	0.642	0.09	0.080
Z_NICG	0.619**	0.094	0.781	0.262	0.776	0.19	0.579	0.022	0.736*	0.217	0.671	0.119	0.155
Z_NPN	0.632*	0.107	0.803(B)	**0.284**	0.772	0.186	0.603	0.046	0.783(B)	**0.264**	0.696(B)	**0.144**	0.165
Z_QN	0.584**	0.059	0.752	0.233	0.747	0.161	0.572	0.015	0.702	0.183	0.667	0.115	0.138
Z_QNZ	0.619**	0.094	0.721	0.202	0.721	0.135	0.583	0.026	0.731	0.212	0.622	0.07	0.115
AUC													
Z_Original	0.531**(W)	Ref	0.500***(W)	Ref	0.568**(W)	Ref	0.619(W)	Ref	0.588*(W)	Ref	0.572(W)	Ref	Ref
Z_Raw	0.835(B)	**0.304**	0.867	0.367	0.854	0.286	0.699	0.08	0.859	0.271	0.692	0.12	**0.279**
Z_Binary	0.720	0.189	0.776	0.276	0.734	0.166	0.679	0.06	0.735	0.147	0.686	0.114	0.157
Z_NICG	0.774	0.243	0.875	0.375	0.873	0.305	0.673	0.054	0.852	0.264	0.726	0.154	0.254
Z_NPN	0.777	0.246	0.887(B)	**0.387**	0.878(B)	**0.31**	0.718	0.099	0.884(B)	**0.296**	0.771(B)	**0.199**	0.271
Z_QN	0.706	0.175	0.803	0.303	0.799	0.231	0.698	0.079	0.826	0.238	0.751	0.179	0.205
Z_QNZ	0.772	0.241	0.801	0.301	0.811	0.243	0.724(B)	**0.105**	0.814	0.226	0.731	0.159	0.234

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ compared with Z_Raw. AUC, area under the curve of the receiver operating characteristic curve; B, best in the corresponding column; BA, balanced accuracy; Delta, difference between the model and Z_Original, best highlighted in bold; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; median of delta, median of the delta by row (normalization method), best highlighted in bold; MLP, Multilayer Perceptron; Ref, reference; RF, Random Forest; SVM_W, (linear) support vector machine with weighting; W, worse in the corresponding column; XGB_W, extreme gradient boosting with weighting; Z_Binary, binarization applied to Z_Raw data; Z_NICG, Normalization using Internal Control Genes (NICG) applied to Z_Raw data; Z_NPN, Non-Parametric Normalization (NPN) applied to Z_Raw data; Z_Original, Z-transformed RNA-seq data in FPKM format, including all gene features shared between the two cohorts; Z_QN, Quantile Normalization (QN) applied to Z_Raw data; Z_QNZ, Quantile Normalization with Z-Score (QNZ) applied to Z_Raw data; Z_Raw, Z_Original data restricted to the selected DEGs.

Table 3. Cross-dataset testing of death classification on all data in three cancer types

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
BA													
Lung adenocarcinoma: tested in OncoSG data (n = 181) with models trained on TCGA (n = 510)													
Z_Original	0.506*** (W)	Ref	0.507*** (W)	Ref	0.520*** (W)	Ref	0.508** (W)	Ref	0.527*** (W)	Ref	0.539*** (W)	Ref	Ref
Z_Raw	0.553	0.047	0.555 (B)	**0.048**	0.574	0.054	0.553 (B)	**0.045**	0.608	0.081	0.611	0.072	**0.051**
Z_Binary	0.520**	0.014	0.524*** (B)	0.017	0.589 (B)	**0.069**	0.524*	0.016	0.645 (B)	**0.118**	0.618 (B)	**0.079**	0.043
Z_NICG	0.557 (B)	**0.051**	0.552	0.045	0.561	0.041	0.536	0.028	0.598	0.071	0.583*	0.044	0.045
Z_NPN	0.547	0.041	0.532	0.025	0.544	0.024	0.539	0.031	0.580	0.053	0.591*	0.052	0.036
Z_QN	0.538	0.032	0.534	0.027	0.562	0.042	0.535	0.027	0.597	0.07	0.591	0.052	0.037
Z_QNZ	0.530	0.024	0.534	0.027	0.557	0.037	0.530	0.022	0.589	0.062	0.602	0.063	0.032
AUC													
Z_Original	0.620	Ref	0.639	Ref	0.525 (W)	Ref	0.648	Ref	0.560** (W)	Ref	0.574*** (W)	Ref	Ref
Z_Raw	0.680	0.06	0.676	0.037	0.599	0.074	0.670	0.022	0.614	0.054	0.658	0.084	0.057
Z_Binary	0.685 (B)	**0.065**	0.677 (B)	**0.038**	0.632	0.107	0.690 (B)	**0.042**	0.654 (B)	**0.094**	0.666	0.092	**0.079**
Z_NICG	0.593 (W)	−0.03	0.608 (W)	−0.031	0.596	0.071	0.602 (W)	−0.05	0.654 (B)	**0.094**	0.647	0.073	0.022
Z_NPN	0.653	0.033	0.658	0.019	0.602	0.077	0.662	0.014	0.627	0.067	0.649	0.075	0.050
Z_QN	0.679	0.059	0.677 (B)	**0.038**	0.633 (B)	**0.108**	0.678	0.03	0.613	0.053	0.681 (B)	**0.107**	0.056
Z_QNZ	0.668	0.048	0.666	0.027	0.612	0.087	0.665	0.017	0.613	0.053	0.667	0.093	0.051
BA													
Melanoma: tested in DFCI data (n = 40) with models trained on TCGA data (n = 360)													
Z_Original	0.587**	Ref	0.574 (W)	Ref	0.536*	Ref	0.587 (W)	Ref	0.509* (W)	Ref	0.545 (W)	Ref	Ref
Z_Raw	0.647 (B)	**0.060**	0.616	0.042	0.583	0.047	0.639	0.052	0.577	0.068	0.613	0.068	0.056

(continued)

Table 3. *(continued)*

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_Binary	0.628	0.041	0.641 (B)	**0.067**	0.585	0.049	0.648	0.061	0.613 (B)	**0.104**	0.614	0.069	**0.064**
Z_NICG	0.528** (W)	−0.059	0.580	0.006	0.541*	0.005	0.639	0.052	0.544	0.035	0.602	0.057	0.021
Z_NPN	0.608	0.021	0.636	0.062	0.595* (B)	**0.059**	0.620	0.033	0.583	0.074	0.617	0.072	0.061
Z_QN	0.537***	−0.050	0.612	0.038	0.526* (W)	−0.010	0.650 (B)	**0.063**	0.558	0.049	0.620 (B)	**0.075**	0.044
Z_QNZ	0.578**	−0.009	0.576	0.002	0.546	0.010	0.622	0.035	0.546	0.037	0.600	0.055	0.023
AUC													
Z_Original	0.616**	Ref	0.593	Ref	0.591	Ref	0.614 (W)	Ref	0.621	Ref	0.621	Ref	Ref
Z_Raw	0.664 (B)	**0.048**	0.640	0.047	0.620 (B)	**0.029**	0.661	0.047	0.661 (B)	**0.040**	0.671 (B)	**0.050**	**0.047**
Z_Binary	0.647	0.031	0.652 (B)	**0.059**	0.555**	−0.036	0.667	0.053	0.646	0.025	0.625	0.004	0.028
Z_NICG	0.481*** (W)	−0.135	0.578* (W)	−0.015	0.591	0.000	0.706** (B)	**0.092**	0.543***	−0.078	0.636	0.015	−0.008
Z_NPN	0.611**	−0.005	0.633	0.040	0.613	0.022	0.699*	0.085	0.573**	−0.048	0.615 (W)	−0.006	0.009
Z_QN	0.547	−0.069	0.610	0.017	0.507** (W)	−0.084	0.703	0.089	0.528***	−0.093	0.633	0.012	−0.029
Z_QNZ	0.542*	−0.074	0.596	0.003	0.530	−0.061	0.663	0.049	0.508** (W)	−0.113	0.629	0.008	−0.029
Glioblastoma: death classification tested in CPTAC data (n = 97) with models trained on TCGA data (n = 145)													
BA													
Z_Original	0.638***	Ref	0.629***	Ref	0.619**	Ref	0.516 (W)	Ref	0.551*** (W)	Ref	0.572 (W)	Ref	Ref
Z_Raw	0.650	0.012	0.654 (B)	0.025	0.580	−0.039	0.558	0.042	0.637	0.086	0.610	0.038	0.032

(continued)

Table 3. *(continued)*

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_Binary	0.578***	−0.060	0.610* (W)	−0.019	0.579 (W)	−0.040	0.549	0.033	0.578*	0.027	0.614	0.042	0.004
Z_NICG	0.621*	−0.017	0.647	0.018	0.661*** (B)	**0.042**	0.581	0.065	0.657 (B)	**0.106**	0.625	0.053	**0.048**
Z_NPN	0.657 (B)	**0.019**	0.638*	0.009	0.647*	0.028	0.575	0.059	0.642	0.091	0.634 (B)	**0.062**	0.044
Z_QN	0.568*** (W)	−0.070	0.654* (B)	**0.025**	0.617	−0.002	0.627 (B)	**0.111**	0.638	0.087	0.626	0.054	0.040
Z_QNZ	0.614*** (W)	−0.024	0.630**	0.001	0.633	0.014	0.613	0.097	0.626*	0.075	0.620	0.048	0.031
AUC													
Z_Original	0.710**	Ref	0.696***	Ref	0.684***	Ref	0.642* (W)	Ref	0.727** (B)	**Ref**	0.692	Ref	Ref
Z_Raw	0.787 (B)	**0.077**	0.703	0.007	0.659	−0.025	0.693	0.051	0.689	−0.038	0.687	−0.005	0.001
Z_Binary	0.667	−0.043	0.690	−0.006	0.642 (W)	−0.042	0.682	0.040	0.643 (W)	−0.084	0.707	0.015	−0.024
Z_NICG	0.666 (W)	−0.044	0.695	−0.001	0.745*** (B)	**0.061**	0.659	0.017	0.702	−0.025	0.648 (W)	−0.044	−0.013
Z_NPN	0.738	0.028	0.713** (B)	**0.017**	0.707***	0.023	0.698*	0.056	0.687	−0.040	0.731** (B)	**0.039**	**0.026**
Z_QN	0.744	0.034	0.695*	−0.001	0.682	−0.002	0.713* (B)	**0.071**	0.679**	−0.048	0.694**	0.002	0.001
Z_QNZ	0.704	−0.006	0.659* (W)	−0.037	0.701	0.045	0.687*	0.045	0.685*	−0.042	0.677	−0.015	−0.011

*P < 0.05; **P < 0.01; ***P < 0.001 compared with Z_Raw. AUC, area under the curve of the receiver operating characteristic curve; B, best in the corresponding column; BA, balanced accuracy; Delta, difference between the model and Z_Original, best highlighted in bold; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; median of delta, median of the delta by row (normalization method), best highlighted in bold; MLP, Multilayer Perceptron; Ref, reference; RF, Random Forest; SVM_W, (linear) support vector machine with weighting; W, worse in the corresponding column; XGB_W, extreme gradient boosting with weighting; Z_Binary, binarization applied to Z_Raw data; Z_NICG, Normalization using Internal Control Genes (NICG) applied to Z_Raw data; Z_NPN, Non-Parametric Normalization (NPN) applied to Z_Raw data; Z_Original, Z-transformed RNA-seq data in FPKM format, including all gene features shared between the two cohorts; Z_QN, Quantile Normalization (QN) applied to Z_Raw data; Z_QNZ, Quantile Normalization with Z-Score (QNZ) applied to Z_Raw data; Z_Raw, Z_Original data restricted to the selected DEGs.

Table 4. Intra-dataset testing of death classification on molecular data alone in three cancer types

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
BA													
						Lung adenocarcinoma							
Z_Original	0.622**	Ref	0.619(W)	Ref	0.606**(W)	Ref	0.500**(W)	Ref	0.649(W)	Ref	0.577(W)	Ref	Ref
Z_Raw	0.781(B)	**0.159**	0.829	0.21	0.768	0.162	0.563(B)	**0.063**	0.834(B)	**0.185**	0.739(B)	0.162	**0.162**
Z_Binary	0.760	0.138	0.767	0.148	0.731	0.125	0.521	0.021	0.756	0.107	0.685	0.108	0.117
Z_NICG	0.580***(W)	−0.042	0.835	0.216	0.848(B)	**0.242**	0.563(B)	**0.063**	0.821	0.172	0.682	0.105	0.139
Z_NPN	0.635**	0.013	0.838(B)	**0.219**	0.830	0.224	0.542	0.042	0.794	0.145	0.689	0.112	0.129
Z_QN	0.664**	0.042	0.801	0.182	0.741	0.135	0.563(B)	**0.063**	0.816	0.167	0.672	0.095	0.115
Z_QNZ	0.630**	0.008	0.794	0.175	0.757	0.151	0.542	0.042	0.780	0.131	0.702	0.125	0.128
AUC													
Z_Original	0.641*(W)	Ref	0.643*(W)	Ref	0.685(W)	Ref	0.625(W)	Ref	0.694(W)	Ref	0.617(W)	Ref	Ref
Z_Raw	0.923(B)	**0.282**	0.935(B)	**0.292**	0.915	0.23	0.808	0.183	0.895	0.201	0.875(B)	0.258	**0.244**
Z_Binary	0.835	0.194	0.827**	0.184	0.815	0.13	0.815	0.19	0.853	0.159	0.825	0.208	0.187
Z_NICG	0.821	0.18	0.895	0.252	0.925(B)	**0.24**	0.764	0.139	0.907(B)	**0.213**	0.786	0.169	0.197
Z_NPN	0.875	0.234	0.915	0.272	0.919	0.234	0.815	0.19	0.899	0.205	0.825	0.208	0.221
Z_QN	0.861*	0.22	0.899	0.256	0.871	0.186	0.810	0.185	0.893	0.199	0.837	0.22	0.210
Z_QNZ	0.885	0.244	0.883	0.24	0.857	0.172	0.839(B)	**0.214**	0.859	0.165	0.839	0.222	0.218
						Melanoma							
BA													
Z_Original	0.592(W)	Ref	0.576(W)	Ref	0.580(W)	Ref	0.574(W)	Ref	0.648(W)	Ref	0.605(W)	Ref	Ref
Z_Raw	0.720	0.128	0.716	0.14	0.708	0.128	0.692(B)	**0.118**	0.695	0.047	0.675	0.07	**0.123**
Z_Binary	0.685	0.093	0.690	0.114	0.684	0.104	0.676	0.102	0.705	0.057	0.661	0.056	0.098
Z_NICG	0.669	0.077	0.712	0.136	0.691	0.111	0.690	0.116	0.706	0.058	0.697(B)	**0.092**	0.102
Z_NPN	0.675	0.083	0.707	0.131	0.712	0.132	0.679	0.105	0.702	0.054	0.683	0.078	0.094
Z_QN	0.717	0.125	0.718	0.142	0.723(B)	**0.143**	0.680	0.106	0.716	0.068	0.697(B)	**0.092**	0.116
Z_QNZ	0.724(B)	**0.132**	0.720(B)	**0.144**	0.721	0.141	0.687	0.113	0.732(B)	**0.084**	0.677	0.072	**0.123**
AUC													

(continued)

Table 4. *(continued)*

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_Original	0.633(W)	Ref	0.610(W)	Ref	0.638(W)	Ref	0.627(W)	Ref	0.672(W)	Ref	0.653(W)	Ref	Ref
Z_Raw	0.802(B)	**0.169**	0.787(B)	**0.177**	0.769	0.131	0.743	0.116	0.773	0.101	0.736	0.083	0.124
Z_Binary	0.759	0.126	0.766	0.156	0.741	0.103	0.731	0.104	0.763	0.091	0.713	0.06	0.104
Z_NICG	0.716	0.083	0.781	0.171	0.760	0.122	0.713	0.086	0.777	0.105	0.765(B)	**0.112**	0.109
Z_NPN	0.741	0.108	0.782	0.172	0.786	0.148	0.741	0.114	0.783	0.111	0.744	0.091	0.113
Z_QN	0.789	0.156	0.783	0.173	0.809(B)	**0.171**	0.733	0.106	0.781	0.109	0.758	0.105	0.133
Z_QNZ	0.799	0.166	0.786	0.176	0.804	0.166	0.744(B)	**0.117**	0.784(B)	**0.112**	0.757	0.104	**0.142**
BA													
Z_Original	0.570**	Ref	0.515**(W)	Ref	0.547*(W)	Ref	0.539(W)	Ref	0.526**(W)	Ref	0.524(W)	Ref	Ref
Z_Raw	0.776(B)	**0.206**	0.746(B)	**0.231**	0.760(B)	**0.213**	0.592	0.053	0.750(B)	**0.224**	0.632	0.108	**0.210**
Z_Binary	0.731	0.161	0.690	0.175	0.718	0.171	0.590	0.051	0.745	0.219	0.668(B)	**0.144**	0.166
Z_NICG	0.615**	0.045	0.724	0.209	0.726	0.179	0.584	0.045	0.741	0.215	0.643	0.119	0.149
Z_NPN	0.670*	0.1	0.704	0.189	0.693	0.146	0.624(B)	**0.085**	0.720	0.194	0.636	0.112	0.129
Z_QN	0.564**(W)	−0.006	0.693	0.178	0.685	0.138	0.580	0.041	0.666	0.14	0.623	0.099	0.119
Z_QNZ	0.581**	0.011	0.696	0.181	0.694	0.147	0.566	0.027	0.697	0.171	0.604	0.08	0.114
AUC													
Z_Original	0.577***(W)	Ref	0.542**(W)	Ref	0.649*(W)	Ref	0.675(W)	Ref	0.570**(W)	Ref	0.566**(W)	Ref	Ref
Z_Raw	0.853(B)	**0.276**	0.842	0.3	0.848	0.199	0.710	0.035	0.855	0.285	0.730	0.164	**0.238**
Z_Binary	0.846	0.269	0.801	0.259	0.782	0.133	0.767(B)	**0.092**	0.803	0.233	0.797(B)	**0.231**	0.232
Z_NICG	0.761*	0.184	0.847(B)	**0.305**	0.863(B)	**0.214**	0.688	0.013	0.865(B)	**0.295**	0.695	0.129	0.199
Z_NPN	0.763*	0.186	0.845	0.303	0.825	0.176	0.712	0.037	0.844	0.274	0.730	0.164	0.181
Z_QN	0.657**	0.08	0.775*	0.233	0.779	0.13	0.686	0.011	0.780*	0.21	0.666	0.1	0.115
Z_QNZ	0.653**	0.076	0.816	0.274	0.748	0.099	0.676	0.001	0.825	0.255	0.717	0.151	0.125

Glioblastoma

*P < 0.05; **P < 0.01; ***P < 0.001 compared with Z_Raw. AUC, area under the curve of the receiver operating characteristic curve; B, best in the corresponding column; BA, balanced accuracy; Delta, difference between the model and Z_Original, best highlighted in bold; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; median of delta, median of the delta by row (normalization method), best highlighted in bold; MLP, Multilayer Perceptron; Ref, reference; RF, Random Forest; SVM_W, (linear) support vector machine with weighting; W, worse in the corresponding column; XGB_W, extreme gradient boosting with weighting; Z_Binary, binarization applied to Z_Raw data; Z_NICG, Normalization using Internal Control Genes (NICG) applied to Z_Raw data; Z_NPN, Non-Parametric Normalization (NPN) applied to Z_Raw data; Z_Original, Z-transformed RNA-seq data in FPKM format, including all gene features shared between the two cohorts; Z_QN, Quantile Normalization (QN) applied to Z_Raw data; Z_QNZ, Quantile Normalization with Z-Score (QNZ) applied to Z_Raw data; Z_Raw, Z_Original data restricted to the selected DEGs.

Table 5. Cross-dataset testing of death classification on molecular data alone in three cancer types

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
BA													
Lung adenocarcinoma: death classification trained in TCGA (n = 510) and tested in OncoSG (n = 181) dataset													
Z_Original	0.539***	Ref	0.513***	Ref	0.527*	Ref	0.515***	Ref	0.586***	Ref	0.532***	Ref	Ref
Z_Raw	0.613 (B)	**0.074**	0.566	0.053	0.569	0.042	0.557	0.042	0.630 (B)	0.044	0.608 (B)	**0.076**	**0.049**
Z_Binary	0.500*** (W)	−0.039	0.500*** (W)	−0.01	0.500*** (W)	−0.03	0.500*** (W)	−0.02	0.500*** (W)	−0.086	0.500*** (W)	−0.032	−0.030
Z_NICG	0.537***	−0.002	0.657** (B)*	**0.144**	0.579 (B)	**0.052**	0.559 (B)	**0.044**	0.630 (B)	**0.044**	0.604	0.072	0.048
Z_NPN	0.578**	0.039	0.634***	0.121	0.529**	0.002	0.541*	0.026	0.621	0.035	0.578	0.046	0.037
Z_QN	0.549**	0.01	0.583	0.07	0.55	0.023	0.537*	0.022	0.610**	0.024	0.590	0.058	0.024
Z_QNZ	0.557***	0.018	0.577	0.064	0.566	0.039	0.529***	0.014	0.628	0.042	0.583	0.051	0.041
AUC													
Z_Original	0.553***	Ref	0.531**	Ref	0.537*	Ref	0.629***	Ref	0.601***	Ref	0.584***	Ref	Ref
Z_Raw	0.633	0.08	0.564	0.033	0.567	0.03	0.701 (B)	**0.072**	0.661	0.06	0.655 (B)	**0.071**	0.066
Z_Binary	0.500*** (W)	−0.053	0.500** (W)	−0.03	0.500 (W)	−0.04	0.500*** (W)	−0.13	0.500*** (W)	−0.101	0.500*** (W)	−0.084	−0.069
Z_NICG	0.601*	0.048	0.687*** (B)	**0.156**	0.589*	0.052	0.686*	0.057	0.684 (B)	**0.083**	0.647	0.060	0.060
Z_NPN	0.645 (B)	**0.092**	0.647**	0.116	0.634* (B)	**0.097**	0.662**	0.033	0.643	0.042	0.608**	0.024	0.067
Z_QN	0.630	0.077	0.620**	0.089	0.625**	0.088	0.667	0.038	0.651	0.05	0.650	0.066	**0.072**
Z_QNZ	0.626	0.073	0.622**	0.091	0.618**	0.081	0.684*	0.055	0.642	0.041	0.622	0.038	0.064
Melanoma: death classification tested in DFCI data (n = 40) with models trained on TCGA data (n = 360)													
BA													
Z_Original	0.582*	Ref	0.594* (W)	Ref	0.523 (W)	Ref	0.620	Ref	0.500* (W)	Ref	0.560** (W)	Ref	Ref
Z_Raw	0.616	0.034	0.594 (W)	0	0.553	0.03	0.634	0.014	0.626	0.126	0.632 (B)	**0.072**	0.032
Z_Binary	0.564* (W)	−0.018	0.637* (B)	**0.043**	0.575	0.052	0.631	0.011	0.654 (B)	**0.154**	0.595	0.035	0.039
Z_NICG	0.587	0.005	0.630	0.036	0.543	0.02	0.607	−0.01	0.637	0.137	0.563**	0.003	0.013

(continued)

Table 5. *(continued)*

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_NPN	0.650 (B)	**0.068**	0.617	0.023	0.582 (B)	**0.059**	0.617	−0.003	0.631	0.131	0.584	0.024	**0.041**
Z_QN	0.600	0.018	0.617*	0.023	0.582 (B)	**0.059**	0.661 (B)	**0.041**	0.594*	0.094	0.590*	0.03	0.036
Z_QNZ	0.572*	−0.01	0.636	0.042	0.561	0.038	0.602 (W)	−0.02	0.591*	0.091	0.611	0.051	0.040
AUC													
Z_Original	0.574*	Ref	0.579** (W)	Ref	0.492 (W)	Ref	0.647	Ref	0.544** (W)	Ref	0.581	Ref	Ref
Z_Raw	0.607	0.033	0.579 (W)	0	0.570	0.078	0.682	0.035	0.667 (B)	**0.123**	0.651 (B)	**0.07**	**0.053**
Z_Binary	0.544* (W)	−0.03	0.628*	0.049	0.548	0.056	0.655	0.008	0.656	0.112	0.580*	−0.001	0.029
Z_NICG	0.624 (B)	**0.05**	0.592**	0.013	0.629 (B)	**0.137**	0.626 (W)	−0.02	0.646	0.102	0.563*** (W)	−0.018	0.032
Z_NPN	0.610	0.036	0.617**	0.038	0.583	0.091	0.651	0.004	0.603***	0.059	0.601	0.02	0.037
Z_QN	0.589	0.015	0.641** (B)	**0.062**	0.527	0.035	0.686 (B)	**0.039**	0.605** (B)	0.061	0.619	0.038	0.039
Z_QNZ	0.592	0.018	0.638**	0.059	0.549	0.057	0.650	0.003	0.593**	0.049	0.639	0.058	**0.053**

Glioblastoma: death classification tested in CPTAC data (n = 97) with models trained on TCGA data (n = 145)

BA

Normalization method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_Original	0.602	Ref	0.548*** (W)	Ref	0.516** (W)	Ref	0.527**	Ref	0.536*** (W)	Ref	0.529 (W)	Ref	Ref
Z_Raw	0.620 (B)	**0.018**	0.599 (B)	**0.051**	0.558	0.042	0.529	0.002	0.592	0.056	0.555	0.026	0.034
Z_Binary	0.564***	−0.038	0.580*	0.032	0.574	0.058	0.518 (W)	−0.01	0.552	0.016	0.577	0.048	0.024
Z_NICG	0.567***	−0.035	0.594	0.046	0.604** (B)	**0.088**	0.542	0.015	0.604 (B)	**0.068**	0.563	0.034	0.040
Z_NPN	0.610	0.008	0.598	0.05	0.590	0.074	0.529	0.002	0.600	0.064	0.571	0.042	**0.046**
Z_QN	0.538*** (W)	−0.064	0.583	0.035	0.571	0.055	0.539	0.012	0.586	0.05	0.561	0.032	0.034
Z_QNZ	0.546***	−0.056	0.574	0.026	0.578	0.062	0.545 (B)	**0.018**	0.573	0.037	0.589 (B)	**0.06**	0.031
AUC													

(continued)

Table 5. (continued)

Normaliza-tion method	LASSO	Delta	LR	Delta	MLP	Delta	RF	Delta	SVM_W	Delta	XGB_W	Delta	Median of delta
Z_Original	0.599** (W)	Ref	0.559 (W)	Ref	0.564*** (W)	Ref	0.579*** (W)	Ref	0.63**	Ref	0.571*** (W)	Ref	Ref
Z_Raw	0.636	0.037	0.637	0.078	0.593	0.029	0.633 (B)	**0.054**	0.651 (B)	**0.021**	0.631	0.06	0.046
Z_Binary	0.648	0.049	0.639	0.08	0.619***	0.055	0.616	0.037	0.588** (W)	−0.042	0.633	0.062	0.052
Z_NICG	0.613	0.014	0.634	0.075	0.631***	0.067	0.602	0.023	0.622**	−0.008	0.646 (B)	**0.075**	0.045
Z_NPN	0.663 (B)	**0.064**	0.640 (B)	**0.081**	0.655*** (B)	**0.091**	0.622	0.043	0.651 (B)	**0.021**	0.626	0.021	**0.060**
Z_QN	0.600	0.001	0.605	0.046	0.603	0.039	0.581*	0.002	0.628	−0.002	0.617	−0.002	0.021
Z_QNZ	0.601	0.002	0.618	0.059	0.632**	0.068	0.607	0.028	0.622**	−0.008	0.601	−0.008	0.029

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ compared with Z_Raw. AUC, area under the curve of the receiver operating characteristic curve; B, best in the corresponding column; BA, balanced accuracy; Delta, differ-ence between the model and Z_Original, best highlighted in bold; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; median of delta, median of the delta by row (normalization method), best highlighted in bold; MLP, Multilayer Perceptron; Ref, reference; RF, Random Forest; SVM_W, (linear) support vector machine with weighting; W, worse in the corresponding column; XGB_W, extreme gradient boosting with weighting; Z_Binary, binarization applied to Z_Raw data; Z_NICG, Normalization using Internal Control Genes (NICG) applied to Z_Raw data; Z_NPN, Non-Parametric Normalization (NPN) applied to Z_Raw data; Z_Original, Z-transformed RNA-seq data in FPKM format, including all gene features shared between the two cohorts; Z_QN, Quantile Normalization (QN) applied to Z_QNZ, Quantile Normalization with Z-Score (QNZ) applied to Z_Raw data; Z_Raw, Z_Original data restricted to the selected DEGs.

Our analyses were conducted on a specific set of transcriptomic and clinical datasets for each of the three cancer types, albeit repeated in three pairs of cancer datasets, and on a selected repertoire of ML models and preprocessing techniques. Future work is required to examine the generalizability of our specific quantitative findings to other datasets or ML methods, such as those for diabetes, digestive diseases, and neurological diseases. Moreover, the consistently good performance of LASSO is an empirical finding and warrants additional theoretical and experimental research. It will also be interesting to assess whether and how other regularized methods, as well as batch-effect correction strategies, can mitigate overfitting yet maintain cross-dataset performance. Future work will extend the current binary survival prediction to time-to-event survival analyses to leverage follow-up information. Finally, developing novel evaluation metrics that better capture cross-dataset robustness will be highly useful but is beyond this study's scope and was not performed.

Conclusions

Our findings challenge the reliance on normalization and intra-dataset evaluation, advocating for regularized models and cross-dataset validation to improve the generalizability of ML modeling. Future work should explore optimal preprocessing strategies for specific data types and develop standardized cross-dataset evaluation frameworks to advance bioinformatics ML applications.

Acknowledgments

None.

Conflict of interest

Lanjing Zhang is a deputy editor-in-chief of *Journal of Clinical and Translational Pathology*. The authors declare no other conflicts of interest.

Author contributions

Study conceptualization and design, ensuring data access, accuracy and integrity (LZ), and manuscript writing (FD and LZ). Both authors contributed to the writing or revision of the review article and approved the final publication version.

Ethical statement

This exempt study using publicly available de-identified data did not require IRB review. Data acquisition and use complied with cBioPortal's data access policies and ethical guidelines. All procedures were conducted in accordance with the principles of the Declaration of Helsinki (as revised in 2024).

Data sharing statement

The datasets used in this study are available on the cBio-

Portal website (https://www.cbioportal.org/). The program code is available from the corresponding author on reasonable request.

References

[1] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286(5439):531–537. doi:10.1126/science.286.5439.531, PMID:10521349.

[2] Deng F, Zhao L, Yu N, Lin Y, Zhang L. Union With Recursive Feature Elimination: A Feature Selection Framework to Improve the Classification Performance of Multicategory Causes of Death in Colorectal Cancer. Lab Invest 2024;104(3):100320. doi:10.1016/j.labinv.2023.100320, PMID:38158124.

[3] Deng F, Zhou H, Lin Y, Heim JA, Shen L, Li Y, et al. Predict multicategory causes of death in lung cancer patients using clinicopathologic factors. Comput Biol Med 2021;129:104161. doi:10.1016/j.compbiomed.2020.104161, PMID:33307409.

[4] Deng F, Shen L, Wang H, Zhang L. Classify multicategory outcome in patients with lung adenocarcinoma using clinical, transcriptomic and clinico-transcriptomic data: machine learning versus multinomial models. Am J Cancer Res 2020;10(12):4624–4639. doi:10.13140/RG.2.2.15071.48802/1.

[5] Cui M, Deng F, Disis ML, Cheng C, Zhang L. Advances in the Clinical Application of High-throughput Proteomics. Explor Res Hypothesis Med 2024;9(3):209–220. doi:10.14218/erhm.2024.00006, PMID:39148720.

[6] Cui M, Cheng C, Zhang L. High-throughput proteomics: a methodological mini-review. Lab Invest 2022;102(11):1170–1181. doi:10.1038/s41374-022-00830-7, PMID:35922478.

[7] Liu DD, Zhang L. Trends in the characteristics of human functional genomic data on the gene expression omnibus, 2001-2017. Lab Invest 2019;99(1):118–127. doi:10.1038/s41374-018-0125-5, PMID:30206311.

[8] Bernau C, Riester M, Boulesteix AL, Parmigiani G, Huttenhower C, Waldron L, et al. Cross-study validation for the assessment of prediction algorithms. Bioinformatics 2014;30(12):i105–i112. doi:10.1093/bioinformatics/btu279, PMID:24931973.

[9] Cohen-Davidi E, Veksler-Lublinsky I. Benchmarking the negatives: Effect of negative data generation on the classification of miRNA-mRNA interactions. PLoS Comput Biol 2024;20(8):e1012385. doi:10.1371/journal.pcbi.1012385, PMID:39186797.

[10] Mohammadzadeh-Vardin T, Ghareyazi A, Gharizadeh A, Abbasi K, Rabiee HR. DeepDRA: Drug repurposing using multi-omics data integration with autoencoders. PLoS One 2024;19(7):e0307649. doi:10.1371/journal.pone.0307649, PMID:39058696.

[11] Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiol Artif Intell 2022;4(3):e210064. doi:10.1148/ryai.210064, PMID:35652114.

[12] Feng CH, Deng F, Disis ML, Gao N, Zhang L. Towards machine learning fairness in classifying multicategory causes of deaths in colorectal or lung cancer patients. Brief Bioinform 2025;26(4):bbaf398. doi:10.1093/bib/bbaf398, PMID:40794953.

[13] Deng F, Feng CH, Gao N, Zhang L. Normalization and Selecting Non-Differentially Expressed Genes Improve Machine Learning Modelling of Cross-Platform Transcriptomic Data. Trans Artif Intell 2025;1(1):5. doi:10.53941/tai.2025.100005, PMID:40630982.

[14] Sun R, Zhu H, Wang Y, Wang J, Jiang C, Cao Q, et al. Circular RNA expression and the competitive endogenous RNA network in pathological, age-related macular degeneration events: A cross-platform normalization study. J Biomed Res 2023;37(5):367–381. doi:10.7555/JBR.37.20230010, PMID:37366063.

[15] Foltz SM, Greene CS, Taroni JN. Cross-platform normalization enables machine learning model training on microarray and RNA-seq data simultaneously. Commun Biol 2023;6(1):222. doi:10.1038/s42003-023-04588-6, PMID:36841852.

[16] Koo B, Kim J, Nam Y, Kim Y. The Performance of Post-Fall Detection Using the Cross-Dataset: Feature Vectors, Classifiers and Processing Conditions. Sensors (Basel) 2021;21(14):4638. doi:10.3390/s21144638, PMID:34300378.

[17] Junet V, Farrés J, Mas JM, Daura X. CuBlock: a cross-platform normalization method for gene-expression microarrays. Bioinformatics 2021;37(16):2365–2373. doi:10.1093/bioinformatics/btab105, PMID:33609102.

[18] Montesinos López OA, Montesinos López A, Crossa J. Overfitting, model tuning, and evaluation of prediction performance. Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham: Springer; 2022:109–139. doi:10.1007/978-3-030-89010-0_4.

[19] Krawczuk J, Łukaszuk T. The feature selection bias problem in relation to high-dimensional gene data. Artif Intell Med 2016;66:63–71. doi:10.1016/j.artmed.2015.11.001, PMID:26674595.

[20] Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95(1):14–18. doi:10.1093/jnci/95.1.14, PMID:12509396.

[21] Vujović Ž. Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications 2021;12(6):599–606. doi:10.14569/IJACSA.2021.0120670.

[22] Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. arXiv 2018. Available from: https://doi.org/10.48550/arXiv.1811.12808.

[23] Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002;99(10):6562–6566. doi:10.1073/pnas.102102699, PMID:11983868.

[24] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 2019;20(1):296. doi:10.1186/s13059-019-1874-1, PMID:31870423.

[25] Huang D, Chow T. Effective gene selection method with small sample sets using gradient-based and point injection techniques. IEEE/ACM Trans Comput Biol Bioinform 2007;4(3):467–475. doi:10.1109/tcbb.2007.1021, PMID:17666766.

[26] Xiong Y, Ling QH, Han F, Liu QH. An efficient gene selection method for microarray data based on LASSO and BPSO. BMC Bioinformatics 2019;20(Suppl 22):715. doi:10.1186/s12859-019-3228-0, PMID:31888444.

[27] Smith AM, Walsh JR, Long J, Davis CB, Henstock P, Hodge MR, et al. Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. BMC Bioinformatics 2020;21(1):119. doi:10.1186/s12859-020-3427-8, PMID:32197580.

[28] Deng F, Zhang Y, Zhang L. Toward the Best Generalizable Performance of Machine Learning in Modeling Omic and Clinical Data. Lab Invest 2025;105(12):104253. doi:10.1016/j.labinv.2025.104253, PMID:41106592.

[29] Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

[30] Benchekroun M, Velmovitsky PE, Istrate D, Zalc V, Morita PP, Lenne D. Cross Dataset Analysis for Generalizability of HRV-Based Stress Detection Models. Sensors (Basel) 2023;23(4):1807. doi:10.3390/s23041807, PMID:36850407.

[31] Tarazona S, Balzano-Nogueira L, Gómez-Cabrero D, Schmidt A, Imhof A, Hankemeier T, et al. Harmonization of quality metrics and power calculation in multi-omic studies. Nat Commun 2020;11(1):3092. doi:10.1038/s41467-020-16937-8, PMID:32555183.

[32] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multi-dimensional cancer genomics data. Cancer Discov 2012;2(5):401–404. doi:10.1158/2159-8290.CD-12-0095, PMID:22588877.

[33] Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet 2020;52(2):177–186. doi:10.1038/s41588-019-0569-6, PMID:32015526.

[34] Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell 2018;173(2):321–337.e10. doi:10.1016/j.cell.2018.03.035, PMID:29625050.

[35] Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 2015;350(6257):207–211. doi:10.1126/science.aad0095, PMID:26359337.

[36] Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 2018;173(2):291–304.e6. doi:10.1016/j.cell.2018.03.022, PMID:29625048.

[37] Wang LB, Karpova A, Gritsenko MA, Kyle JE, Cao S, Li Y, et al. Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell 2021;39(4):509–528.e20. doi:10.1016/j.ccell.2021.01.006, PMID:33577785.

[38] Deng F, Huang J, Yuan X, Cheng C, Zhang L. Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data. Lab Invest 2021;101(4):430–441. doi:10.1038/s41374-020-00525-x, PMID:33574440.

[39] Feng CH, Disis ML, Cheng C, Zhang L. Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models. Lab Invest 2021;doi:10.1038/s41374-021-00662-x, PMID:34537824.

[40] Bhuva DD, Cursons J, Davis MJ. Stable gene expression for normalisation and single-sample scoring. Nucleic Acids Res 2020;48(19):e113. doi:10.1093/nar/gkaa802, PMID:32997146.

[41] Thompson JA, Tan J, Greene CS. Cross-platform normalization of microarray and RNA-seq data for machine learning applications. PeerJ 2016;4:e1621. doi:10.7717/peerj.1621, PMID:26844019.

[42] Brodsky E, Darkhovsky BS. Non-Parametric Statistical Diagnosis: Problems and Methods. Dordrecht: Springer; 2013.

[43] Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3(7):RESEARCH0034. doi:10.1186/gb-2002-3-7-research0034, PMID:12184808.

[44] Karthik S, Sudha M. A survey on machine learning approaches in gene expression classification in modelling computational diagnostic system for complex diseases. Int J Eng Adv Technol 2018;8(2):182–191.

[45] Dunne RA. A statistical approach to neural networks for pattern recognition. John Wiley & Sons; 2007. doi:10.1002/9780470148150.

[46] Ma B, Meng F, Yan G, Yan H, Chai B, Song F. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Comput Biol Med 2020;121:103761. doi:10.1016/j.compbiomed.2020.103761, PMID:32339094.

[47] Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM. Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships. J Chem Inf Model 2016;56(12):2353–2360. doi:10.1021/acs.jcim.6b00591, PMID:27958738.

[48] Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013. doi:10.1002/9781118548387.

[49] Huang YM, Du SX. Weighted support vector machine for classification with

uneven training class sizes. Proceedings of the 2005 International Conference on Machine Learning and Cybernetics; 2005 Aug 18-21; Guangzhou, China. Piscataway (NJ): IEEE; 2005:4365–4369. doi:10.1109/ICMLC.2005.1527706.

[50] Parmar A, Katariya R, Patel V. A review on random forest: an ensemble classifier. In: Hemanth J, Fernando X, Lafata P, Baig Z (eds). International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI 2018); 2018 Aug 7-8; Coimbatore, India. Cham: Springer; 2019:758–763. doi:10.1007/978-3-030-03146-6_86.

[51] Derrick B, Toher D, White P. Why Welch's test is Type I error robust. Quant Methods Psychol 2016;12(1):30–38. doi:10.20982/tqmp.12.1.p030.

[52] Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. Behav Ecol 2006;17(4):688–690. doi:10.1093/beheco/ark016.

[53] Sedgwick P. A comparison of parametric and non-parametric statistical tests. BMJ 2015;350:h2053. doi:10.1136/bmj.h2053, PMID:25888112.

[54] Hollander M, Wolfe DA, Chicken E. Nonparametric statistical methods. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013.

[55] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545–15550. doi:10.1073/pnas.0506580102, PMID:16199517.

[56] Lu S, Ortiz C, Fürth D, Fischer S, Meletis K, Zador A, et al. Assessing the replicability of spatial gene expression using atlas data from the adult mouse brain. PLoS Biol 2021;19(7):e3001341. doi:10.1371/journal.pbio.3001341, PMID:34280183.

[57] Breiman L. Random Forests. Mach Learn 2001;45(1):5–32. doi:10.1023/A:1010933404324.

[58] Ben Or G, Veksler-Lublinsky I. Comprehensive machine-learning-based analysis of microRNA-target interactions reveals variable transferability of interaction rules across species. BMC Bioinformatics 2021;22(1):264. doi:10.1186/s12859-021-04164-x, PMID:34030625.

[59] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York (NY): Springer; 2009. doi:10.1007/978-0-387-84858-7.

[60] Santos CFGD, Papa JP. Avoiding overfitting: a survey on regularization methods for convolutional neural networks. ACM Comput Surv 2022;54(10s):1–25. doi:10.1145/3510413.

[61] Moradi R, Berangi R, Minaei B. A survey of regularization strategies for deep models. Artif Intell Rev 2020;53(6):3947–3986. doi:10.1007/s10462-019-09784-7.

[62] Aliniya P, Nicolescu M, Nicolescu M, Bebis G. Towards Robust Supervised Pectoral Muscle Segmentation in Mammography Images. J Imaging 2024;10(12):331. doi:10.3390/jimaging10120331, PMID:39728227.

[63] Napoli O, Duarte D, Alves P, Soto DHP, de Oliveira HE, Rocha A, et al. A benchmark for domain adaptation and generalization in smartphone-based human activity recognition. Sci Data 2024;11(1):1192. doi:10.1038/s41597-024-03951-4, PMID:39488523.

[64] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47. doi:10.1093/nar/gkv007, PMID:25605792.

[65] Charmpi K, Chokkalingam M, Johnen R, Beyer A. Optimizing network propagation for multi-omics data integration. PLoS Comput Biol 2021;17(11):e1009161. doi:10.1371/journal.pcbi.1009161, PMID:34762640.

[66] Béal J, Montagud A, Traynard P, Barillot E, Calzone L. Personalization of Logical Models With Multi-Omics Data Allows Clinical Stratification of Patients. Front Physiol 2018;9:1965. doi:10.3389/fphys.2018.01965, PMID:30733688.