LEARNING WHAT TO ENCODE: REPRESENTATION SELECTION FOR COMPETITIVE QUANTUM TABULAR CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Establishing the predictive competitiveness of quantum machine learning against strong classical baselines remains a central challenge in tabular classification. We propose learned representation selection (LRS), a hybrid quantum–classical classifier that learns a discrete representation choice for each encoding slot under a fixed quantum input budget. Instead of prescribing a representation mixture, LRS assigns each slot a feature from complementary representations, allowing the encoded input to adapt to the prediction task. A select–freeze–refit procedure separates learning these assignments from training the final predictor: the selection stage learns slot-wise choices, the freeze stage fixes the resulting discrete configuration, and the refit stage trains the predictor with that configuration held constant. The selected features are angle-encoded into a variational quantum circuit, whose expectation values are combined with a classical feature branch. Across twenty tabular datasets, LRS achieves accuracy and macro-F1 comparable to tuned random forests, XGBoost, and LightGBM, with advantages on several datasets. These results support slot-wise discrete representation selection as a practical approach to competitive hybrid quantum classification under a constrained encoding budget.

1 INTRODUCTION

Quantum machine learning (QML) explores how quantum computation can support learning from data. In variational approaches, inputs are encoded into quantum states, transformed by parameterized circuits, and measured to produce features or predictions. A classical optimizer updates the circuit parameters using a task-specific objective, allowing quantum models to be trained within hybrid learning pipelines (Cerezo et al., 2021). Quantum state transformations offer an alternative family of feature maps: interference and entangling operations can shape the functions accessible to a model, while the encoding determines how classical inputs enter that family (Schuld et al., 2021). This flexibility motivates QML as a framework for investigating useful inductive biases alongside established classical architectures.

Realizing this potential requires addressing both computational constraints and predictive performance. Limited qubit counts restrict the available quantum state space, hardware noise constrains circuit depth, and unfavorable optimization landscapes can hinder training (Cerezo et al., 2021). Even in noiseless simulation, expressive circuits do not automatically yield better generalization. Systematic benchmarks have found that classical models can outperform quantum classifiers on the small learning tasks studied (Bowles et al., 2024). Tabular classification is a particularly demanding setting because tree ensembles remain strong competitors across diverse datasets (Grinsztajn et al., 2022). Establishing competitive prediction in this setting therefore calls for careful design of the entire learning pipeline, including the information supplied to the circuit.

A central design problem arises at the interface between a tabular representation and a resource-constrained quantum encoder. The number of available features can greatly exceed the number of coordinates that a chosen encoding scheme accepts. Compression makes these inputs tractable, but its objective need not coincide with the needs of the downstream classifier. Different feature representations can preserve complementary aspects of the data, and committing to a single rep-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

resentation or a fixed concatenation determines in advance which information reaches the model. Searching over representation counts offers some flexibility, but a count alone does not identify which candidate coordinate should occupy each encoding slot. This motivates the question at the center of this work: *how can a hybrid classifier learn what to encode while keeping its input budget fixed?*

We address this question with learned representation selection (LRS), which treats the source assigned to each encoding slot as a learnable discrete choice. Each slot receives one of two candidate coordinates from complementary representations. A global binary mask, shared across samples within a fitted model, selects exactly one coordinate per slot. The mask is optimized jointly with a hybrid predictor through hard forward decisions and surrogate gradients. Consequently, every forward pass respects the prescribed encoding budget, and the learned source allocation can adapt to the classification objective without increasing the number of encoded coordinates. The interface is specified in terms of candidate representations rather than a particular feature extractor.

To separate representation discovery from prediction on stable inputs, we use a select–freeze–refit procedure. The first stage learns the source choices together with the predictor. We then freeze the mask, reinitialize the predictor, and fit it on the resulting fixed representation. The deployed model uses only the frozen mask and the refitted predictor. This procedure turns representation assignment into an explicit part of model learning and provides a clear basis for evaluating source selection, circuit structure, and the encoding budget separately.

Across twenty binary tabular datasets, LRS achieved predictive performance competitive with strong tree ensembles. Ablation studies examined the roles of representation selection, circuit topology, and the qubit budget. Our contributions are threefold:

- **Learning representation assignment under a fixed encoding budget.** We formulate a per-slot source-selection interface that learns discrete choices between complementary candidate representations while preserving the number of encoded coordinates.
- **Separating selection from final predictor fitting.** We develop a select–freeze–refit procedure that discovers a global hard mask and then trains a freshly initialized predictor on the fixed selected representation.
- **Benchmarking predictive performance and architectural choices.** We evaluate LRS against tuned tree ensembles on twenty tabular datasets and examine representation selection, circuit topology, and encoding width through ablation studies. Per-dataset results and paired statistical tests characterize its relative performance.

2 Related Work

QML for supervised learning. QML approaches to classical prediction include trainable quantum circuits and quantum kernel methods. Mitarai et al. (2018) introduced quantum circuit learning, in which a classical optimizer adjusts circuit parameters to fit nonlinear target functions. Havlíček et al. (2019) investigated both variational classification and kernel estimation using quantum feature spaces, illustrating two ways to connect quantum-state representations with supervised prediction. Their predictive utility on practical datasets remains an empirical question: Bowles et al. (2024) benchmarked twelve quantum models and found that standard classical models generally performed better on the classification tasks they considered. These findings motivate evaluating quantum classifiers against strong classical baselines and examining which components contribute to prediction. LRS complements these approaches by learning which representation supplies each encoding slot under a fixed input budget.

Quantum data encoding. Quantum feature maps encode classical inputs into quantum states through input-dependent operations. Their expressive properties depend on the encoding gates as well as the trainable transformations and measurement observables. Schuld et al. (2021) characterize a class of variational quantum models as partial Fourier series: the encoding determines the accessible frequencies, while the remaining circuit structure controls their coefficients.

Learning from multiple representations. Multiple-view learning develops representations from different descriptions of the same examples. Simple concatenation retains all view-specific coordi-

nates, whereas joint representation learning seeks a common space that captures relationships across views. Deep canonical correlation analysis learns nonlinear transformations of two views by maximizing correlation between their outputs (Andrew et al., 2013). This objective encourages shared structure without directly optimizing a downstream classification loss. More generally, combining representations involves two separate decisions: how each view is constructed and how its information is retained in the final predictor. A shared embedding transforms the views into new coordinates, while coordinate selection preserves a subset of the available candidates. The distinction is consequential when an encoder accepts a prescribed number of inputs, because concatenation alone does not enforce that budget.

Differentiable feature selection. Differentiable feature-selection methods learn input subsets jointly with reconstructive or predictive models. Concrete autoencoders use a prescribed number of selector units and a continuous relaxation to select features that reconstruct the original input (Balın et al., 2019). Stochastic gates instead couple probabilistic feature selection to a supervised objective, controlling sparsity through a regularization penalty (Yamada et al., 2020). These methods differ in both their selection objectives and their control of subset size: a fixed number of selectors specifies the selection budget, whereas regularization trades predictive fit against sparsity. Straight-through estimators provide a complementary mechanism for optimizing discrete gates, retaining hard decisions in the forward pass while propagating surrogate gradients during backpropagation (Bengio et al., 2013).

3 Proposed Method

3.1 A fixed-budget representation interface

Let $x \in \mathbb{R}^d$ and $y \in \{0, 1\}$ denote an input and its target. Two feature extractors, followed by coordinate-wise angle maps, yield candidate vectors $a(x), b(x) \in \mathbb{R}^q$. The extractors are fitted before classifier optimization and held fixed during selection. They can embody different information-preservation objectives. Our implemented instance uses principal component analysis (PCA) and neighborhood components analysis (NCA). PCA constructs orthogonal directions that retain input variance, whereas NCA learns a supervised linear transformation that favors same-class neighbors (Goldberger et al., 2004). The two extractors are fitted independently to the standardized input, yielding variance-oriented and label-informed candidate views. NCA coordinates are not ordered by explained variance and are not intrinsically aligned with PCA coordinates. The selection interface does not require such alignment: its slot pairing specifies the available alternatives. Substituting another extractor changes the candidate bank and requires a separate evaluation.

A mask $m \in \{0, 1\}^q$ defines the encoded representation

$$z_j(x; m) = m_j a_j(x) + (1 - m_j) b_j(x), \qquad j = 1, \ldots, q. \tag{1}$$

Exactly one candidate occupies each slot. The selected feature count is q for every mask, and the count from the first view is $\sum_j m_j$. A slot index is a deterministic pairing convention, not an assertion of semantic correspondence between the two extractors. The selector cannot simultaneously retain a_j and b_j, nor can it move a coordinate between slots. It is therefore distinct from choosing any q coordinates from a bank of $2q$ features.

Figure 1 illustrates how the source mask maps the candidate representations to a fixed-width input and how the selected representation is used during fresh refitting.

3.2 Hard selections with surrogate gradients

Let α_j be a learnable source logit and $\tau > 0$ a surrogate temperature. Define

$$p_j = \mathrm{sigmoid}(\alpha_j/\tau), \quad h_j = \mathbf{1}[\alpha_j \geq 0], \quad \widetilde{m}_j = \mathrm{sg}(h_j - p_j) + p_j, \tag{2}$$

where sg stops the gradient. The forward value is exactly h_j, so training never supplies a fractional mixture to the predictor. During differentiation, the gate uses the sigmoid derivative. In particular,

$$\frac{\widetilde{\partial} z_j}{\partial \alpha_j} = (a_j - b_j) \frac{p_j(1 - p_j)}{\tau}. \tag{3}$$

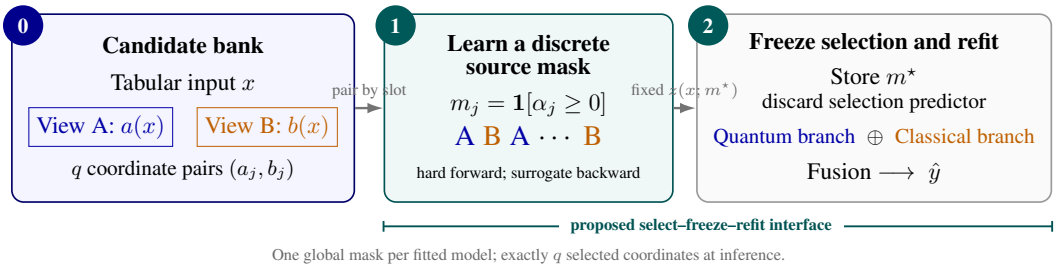


Figure 1: Overview of learned representation selection (LRS). Candidate coordinates are paired by slot, and a global hard mask learns which source occupies each slot. The mask is then frozen and the selection-stage predictor is discarded. A freshly initialized hybrid predictor is refitted on the fixed representation. The frozen representation is supplied to both the quantum and classical branches, whose outputs are fused for prediction.

The tilde emphasizes a surrogate derivative: this is not the true derivative of the discontinuous objective. A coordinate pair with identical values supplies no predictive gate gradient, and a strongly saturated logit supplies little gradient even when the coordinates differ.

The selection-stage objective combines classification and a margin term:

$$\mathcal{L}_{\text{select}}(\theta, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \ell_w(y_i, f_\theta(z(x_i; \widetilde{m}))) + \lambda \frac{1}{q} \sum_{j=1}^{q} \max(0, \mu - |\alpha_j|), \tag{4}$$

where $\ell_w(y, p) = -wy \log p - (1 - y) \log(1 - p)$, θ denotes predictor parameters, and μ is a logit margin. The penalty discourages logits near the hard-decision boundary, but does not favor a particular source count. It is not a cardinality penalty: cardinality is already fixed by Equation 1. A margin term also cannot guarantee a stable or optimal mask. Predictive and regularization gradients can compete, making empirical checks of selection important.

3.3 Freeze the mask, refit the predictor

After joint selection-stage training, we set $m_j^\star = \mathbf{1}[\alpha_j \geq 0]$, discard the fitted predictor parameters, and initialize a fresh predictor. This predictor is optimized on $z(x; m^\star)$ with the classification loss alone. The mask and feature extractors remain fixed. At inference, the stored mask selects coordinates and the refitted predictor produces the probability; the selection-stage predictor is not used.

This procedure removes input switches during the final fitting stage. Algorithm 1 summarizes the select–freeze–refit procedure.

Algorithm 1 Select–freeze–refit procedure. All inputs are restricted to the corresponding training partition.

1. Fit the two feature extractors and coordinate-wise angle maps; construct candidate vectors $a(x_i), b(x_i)$.
2. Initialize source logits α and predictor parameters θ.
3. Jointly optimize Equation 4, using hard forward gates and the surrogate in Equation 2.
4. Freeze $m^\star = \mathbf{1}[\alpha \geq 0]$ and construct $z(x_i; m^\star)$.
5. Reinitialize predictor parameters and fit the classification loss on the fixed selected representation.
6. Return the extractors, angle maps, fixed mask, and refitted predictor.

3.4 Hybrid prediction backbone

The selected angles are encoded through $E(z) = \bigotimes_{j=1}^{q} R_y(s_j z_j)$, where R_y is a single-qubit rotation about the y axis and s_j are bounded, trainable encoding scales. Trainable three-angle single-

qubit rotations (Rot) and ordered controlled-NOT (CNOT) rings form a variational block. The rings operate within two fixed wire groups. If U_θ denotes the resulting circuit and Z_j the Pauli-Z observable on wire j, its output is the vector of local expectations

$$v_j(z) = \langle 0|E(z)^\dagger U_\theta^\dagger Z_j U_\theta E(z)|0\rangle. \tag{5}$$

The classical branch uses a rectified linear unit (ReLU) to compute $c(z) = \mathrm{ReLU}(W_c z + b_c)$. The two branches are concatenated, projected through another ReLU, and read out:

$$f_\theta(z) = \mathrm{sigmoid}\big(w_o^\top \mathrm{ReLU}(W_f[v(z); c(z)] + b_f) + b_o\big)\,. \tag{6}$$

Thus, source selection affects both branches. Improved prediction cannot be assigned to the quantum branch merely because the selector is trained through a hybrid network.

Figure 2 shows the quantum computation and contrasts the two-group topology with the global Ring-CNOT ablation.

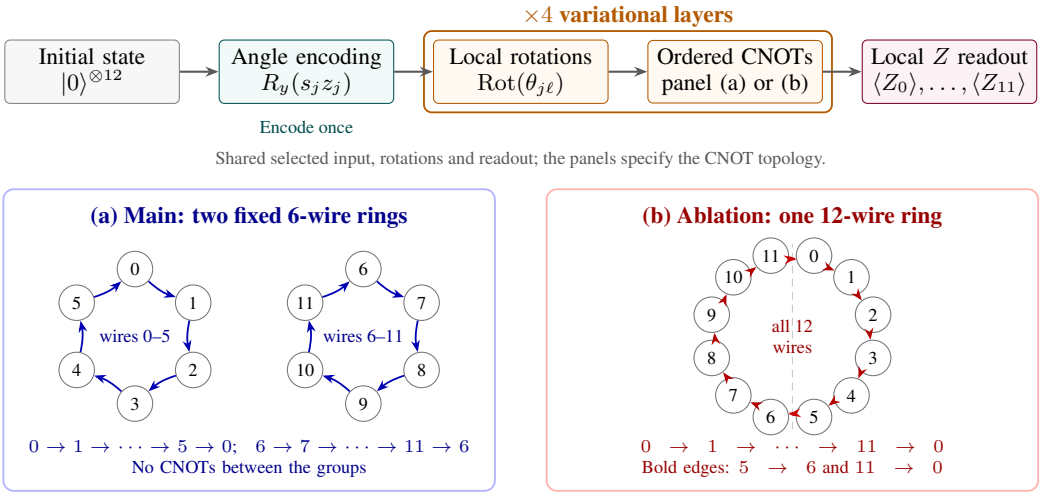


Figure 2: Quantum-branch computation and entangling topology. The selected angles are encoded once, followed by four rotation/CNOT layers and local Pauli-Z readout. Wire groups in (a) are fixed by slot index, not by the selected feature source. Panel (b) replaces the two rings by one global ring; bold edges cross the former group boundary.

3.5 Scope and Structural Properties

The formal mask space contains 2^q binary choices, controlled by q logits. The number of distinct representations may be smaller when coordinates coincide or are padded. A count-allocation baseline instead concatenates a prefix from one extractor with a prefix from the other. These spaces are not nested: a count allocation may keep both members of some pairs and omit other pairs, while LRS always keeps exactly one member of each pair. If the extractors are refitted at different output dimensions, the candidate functions themselves also change. We therefore compare these as alternative allocation strategies rather than presenting learned selection as a strict superset of count tuning.

The circuit ends with pairwise IsingZZ gates. Because those operations commute with the local Z readout, they are observationally redundant for this predictor. Selection and classical fusion can still combine information from coordinates in both groups.

4 EXPERIMENTS

4.1 TASKS, SPLITS, AND MODEL SELECTION

We evaluated twenty binary tabular tasks. The outer evaluation used stratified five-fold cross-validation (CV) with split seed 101. Each method was evaluated on five outer folds per dataset.

For LRS and count allocation, each outer-training partition was split once into stratified 80/20 inner-training/validation subsets. Twenty tree-structured Parzen estimator (TPE) trials selected hyperparameters by validation macro-F1, with binary cross-entropy and trial index as tie breakers. The chosen configuration was retrained on the entire outer-training partition before the held-out outer test fold was scored. Standardization, PCA, NCA, and angle scaling were fitted within the relevant training partition. The selection mask was learned anew during outer refitting; the inner mask was not reused. Test data was not used for hyperparameter or epoch selection within a run.

The evaluated main circuit used twelve wires, four variational layers, and two fixed six-wire rings; throughout the paper, *LRS* without a qualifier denotes this model. Each candidate view contained twelve coordinates, with padding when the effective feature dimension was smaller. Selection and refitting each had a maximum of 50 epochs. Every five epochs, training accuracy was checked; a stage stopped at 99% or higher. This was a training-saturation rule, not validation-based early stopping. The two-stage total could therefore reach 100 epochs per trial and per outer refit. Appendix B specifies all ranges and seeds.

Tree references used 50 TPE trials with three-fold inner CV. All methods used the same outer test indices. The dataset collection had also been used during method development.

4.2 COMPARATORS AND CONTROLLED STUDIES

The tree baselines comprised random forest (RF) (Breiman, 2001), extreme gradient boosting (XGBoost) (Chen & Guestrin, 2016), and Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017). Tables abbreviate XGBoost and LightGBM as XGB and LGBM, respectively; Raw denotes original features. The primary references were RF, XGBoost, and LightGBM on original features, and *Count-12* was a hyperparameter-tuned allocation baseline that kept the total number of encoded features fixed at twelve. It searched over thirteen PCA/NCA allocations, $(p, n) \in \{(0, 12), (1, 11), \ldots, (12, 0)\}$, and selected the allocation with the highest validation macro-F1. For the selected allocation, the model concatenated p PCA coordinates and n NCA coordinates and uploaded them once to the twelve-wire circuit. Changing (p, n) also changed the fitted NCA representation and the boundary between the CNOT-ring groups. In contrast, LRS selected features from a fixed candidate bank while retaining fixed wire groups. Therefore, the comparison reflected differences in feature construction and circuit topology as well as the selection strategy.

The controlled representation ablations comprised a dynamic-group selector, PCA-only 12, NCA-only 12, global Ring-CNOT, and learned selectors with ten and eight wires. Dynamic-group LRS reordered selected PCA coordinates before selected NCA coordinates and placed the ring boundary at the learned source count during fresh refitting. Global Ring-CNOT instead preserved the twelve selected slots and connected all wires in one ordered ring (Figure 2). PCA-only and NCA-only removed the source selector while retaining one twelve-coordinate upload. LRS-10 and LRS-8 retained learned source selection but used fixed 5+5 and 4+4 ring groups. All seven models were evaluated on 20 datasets with five outer folds per dataset. We additionally evaluated trees on the concatenated PCA/NCA bank.

4.3 PREDICTIVE PERFORMANCE AND DATASET DEPENDENCE

LRS achieved 87.15% accuracy and 83.76% macro-F1 on the complete common cohort (Table 1). Count-12 achieved 86.29% and 82.71%; RF achieved 86.73% and 83.61%; XGBoost achieved 85.78% and 82.48%; and LightGBM achieved 86.68% and 83.72%. These small differences in cross-dataset means supported predictive competitiveness in this benchmark.

Table 2 compares all seven models on the complete 20-dataset cohort. The main fixed-6+6 LRS achieved the best average ranks for accuracy (3.03) and macro-F1 (3.18). NCA-only 12 had the highest unweighted means (87.26% and 83.91%), showing that aggregate score and consistency

Table 1: Primary comparison on the 20 complete common datasets. All models shared outer test folds. Dataset means received equal weight. Ranks were computed within these five methods.

Method	Accuracy	Rank	Macro-F1	Rank
LRS	**87.15**	2.80	**83.76**	2.90
Count-12	86.29	3.77	82.71	3.80
RF-Raw	86.73	2.27	83.61	2.25
XGB-Raw	85.78	3.45	82.48	3.45
LGBM-Raw	86.68	2.70	83.72	2.60

across datasets capture different aspects of performance. LRS-10 ranked second in accuracy (3.60), while PCA-only 12 ranked second in macro-F1 (3.60).

Table 2: Representation, topology, and width comparisons on 20 datasets. Each model has 100 outer-test results. Ranks use dataset means across all seven models (lower is better); red, blue, and green mark the top three scores or ranks in each column.

Model	Representation / circuit	Acc.	Rank↓	F1	Rank↓
LRS (main)	Learned source; fixed 6+6 rings	**87.15**	**3.03**	**83.76**	**3.18**
Dynamic-group LRS	Learned source; groups follow selected source count	86.30	4.55	**83.18**	4.43
PCA-only 12	PCA inputs; no source selector	**86.75**	**3.70**	83.01	**3.60**
NCA-only 12	NCA inputs; no source selector	**87.26**	3.90	**83.91**	3.98
Ring-CNOT LRS	Learned source; one 12-wire ring	86.18	4.25	82.84	4.23
LRS-10	Learned source; fixed 5+5 rings	86.39	**3.60**	83.03	**3.70**
LRS-8	Learned source; fixed 4+4 rings	84.54	4.98	80.46	4.90

The cross-dataset conclusion also depended on task composition. Table 3 reports an explicit leave-Hill-Valley-out sensitivity analysis using the same complete cohort. Performance on Hill-Valley appeared particularly sensitive to the PCA/NCA input representation. Excluding this dataset increased the mean performance of the tree ensembles, while LRS remained competitive with these baselines.

Table 3: Descriptive sensitivity: the 19 complete common datasets excluding Hill-Valley. This subset is not the primary benchmark.

Method	Accuracy	Macro-F1
LRS	87.84	84.34
Count-12	86.91	83.19
RF-Raw	**88.28**	85.01
XGB-Raw	87.39	83.92
LGBM-Raw	88.17	**85.06**

Circuit topology. Global Ring-CNOT achieved 86.18% accuracy and 82.84% macro-F1. The main model improved these means by 0.97 and 0.92 percentage points, respectively, and attained higher accuracy on 16 datasets and higher macro-F1 on 15, with one tie for each metric.

Encoding width. LRS-10 achieved 86.39% accuracy and 83.03% macro-F1, trailing the main model by 0.76 and 0.73 points. LRS-8 achieved 84.54% and 80.46%, corresponding to larger gaps of 2.61 and 3.30 points. The ten-wire model therefore retained much of the mean predictive performance with two fewer wires. Increasing the number of qubits expands the encoding budget available to represent the original features. Consistent with this interpretation, mean accuracy and macro-F1 increased as the qubit budget grew from 8 to 10 to 12. Appendix C provides every dataset mean, fold standard deviation, and rank.

Figure 3 compares the distribution of dataset-level mean scores for the primary methods. The broad overlap between LRS and the tree baselines supported their aggregate competitiveness, while the wide dispersion showed that relative performance varied substantially across datasets.

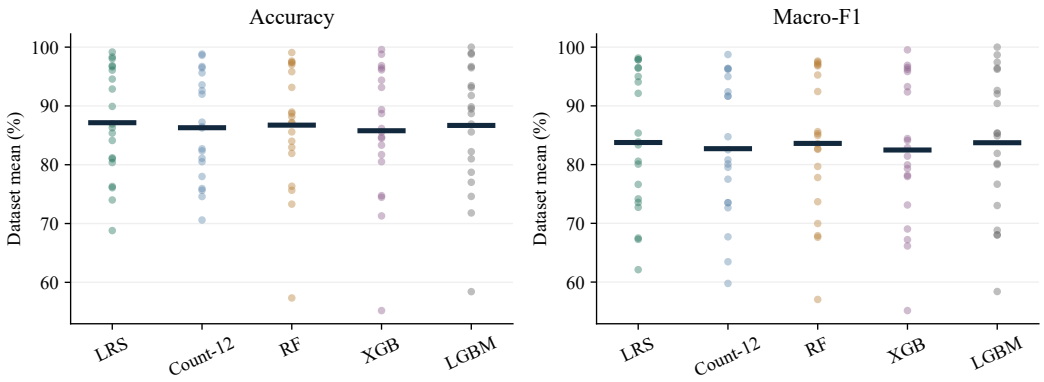


Figure 3: Distribution of complete dataset means for the primary comparison. Each point represents one dataset, not one independent training run; horizontal bars show equal-weight dataset averages. Performance varied substantially across datasets.

Table 4: LRS versus tree ensembles using concatenated PCA12+NCA12 features. Scores are mean accuracy and macro-F1 (%) across 20 datasets, with five outer folds per dataset. Red, blue, and green mark the top three scores in each column.

Model	Input representation	Accuracy	Macro-F1
LRS (main)	12 learned PCA/NCA selections	**87.15**	**83.76**
RF	PCA12+NCA12 (24 features)	**86.83**	**83.53**
XGB	PCA12+NCA12 (24 features)	**86.52**	**83.08**
LGBM	PCA12+NCA12 (24 features)	85.77	82.04

Tree ensembles on PCA/NCA representations. Table 4 compares LRS with tree ensembles trained on the concatenation of twelve PCA and twelve NCA features. LRS achieved the highest mean accuracy and macro-F1, exceeding RF, the strongest tree ensemble in this comparison, by 0.32 and 0.23 percentage points, respectively. Thus, LRS remained competitive when the tree baselines also used PCA/NCA representations, while encoding only twelve selected features. Dataset-level results in Tables 10 and 11 show that LRS outperformed all three trees on Sonar, Colon, Leukemia, and DLBCL, whereas the trees performed substantially better on Hill-Valley.

5 Limitations and Discussion

The method constrains which candidate pairs can compete but does not establish semantic alignment between coordinate systems. Because NCA is supervised, the observed gains may have partly arisen from classical metric learning before the quantum circuit. Indeed, the NCA-only model achieved slightly higher cross-dataset mean scores than the main model, although the paired differences were not statistically significant after multiple-comparison correction. Low-dimensional datasets contained padded slots; therefore, twelve wires did not necessarily represent twelve independent informative coordinates.

The experiments used an exact state-vector simulator rather than noisy hardware. The terminal IsingZZ invariance further limits the effective quantum interaction structure. We make no claims of computational speedup, hardware efficiency, or quantum advantage.

Future work should evaluate a broader range of encoding schemes and feature representations, including autoencoder-based representations. Such comparisons are needed to determine whether the

benefits of learned representation selection generalize beyond the particular representations and angle encoding evaluated here.

6 CONCLUSION

We introduced LRS, a hybrid quantum–classical classifier that learns discrete representation choices under a fixed encoding budget. Hard source selection adapts the model inputs to each task, while freezing the mask before fresh refitting separates representation discovery from predictor training. Across twenty datasets, the fixed-6+6 model achieved competitive predictive performance against tuned tree ensembles and the best average ranks among the seven representation, topology, and width variants. It exceeded global Ring-CNOT and the eight- and ten-wire selectors in mean accuracy and macro-F1, while the ten-wire model retained much of the predictive performance with a smaller register. These results support learned representation selection as a practical interface for allocating limited quantum inputs and motivate extensions to additional feature representations.

Reproducibility Statement

The appendix documents dataset dimensions, partition rules, preprocessing, search spaces, stopping rules, and circuit operations. The representation, topology, and width tables report all 100 outer-fold evaluations for each of the seven models. Tables and paired tests are generated from saved result records, with identical outer-test indices verified across models. The accompanying code includes the fitted configurations, hyperparameter optimization records, and numerical data needed to reproduce the summaries.

AI Use Statement

AI tools were utilized exclusively for language polishing and code debugging. They played no role in generating experimental data or selecting results for publication. The authors have fully verified the final text, code modifications, numerical outcomes, and scientific claims.

References

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1247–1255, 2013. URL https://proceedings.mlr.press/v28/andrew13.html.

Muhammed Fatih Balın, Abubakar Abid, and James Zou. Concrete autoencoders: Differentiable feature selection and reconstruction. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 444–453, 2019. URL https://proceedings.mlr.press/v97/balin19a.html.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013. URL https://arxiv.org/abs/1308.3432.

Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models. *arXiv preprint arXiv:2403.07059*, 2024. URL https://arxiv.org/abs/2403.07059.

Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. doi: 10.1023/A:1010933404324.

M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021. doi: 10.1038/s42254-021-00348-9. URL https://arxiv.org/abs/2012.09265.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. doi: 10.1145/2939672.2939785.

Jacob Goldberger, Geoffrey E. Hinton, Sam T. Roweis, and Ruslan Salakhutdinov. Neighbourhood components analysis. In *Advances in Neural Information Processing Systems*, volume 17, 2004. URL https://papers.nips.cc/paper_files/paper/2004/hash/42fe880812925e520249e808937738d2-Abstract.html.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/0378c7692da36807bdec87ab043cdadc-Abstract-Datasets_and_Benchmarks.html.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567:209–212, 2019. doi: 10.1038/s41586-019-0980-2. URL https://arxiv.org/abs/1804.11326.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98:032309, 2018. doi: 10.1103/PhysRevA.98.032309. URL https://arxiv.org/abs/1803.00745.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103:032430, 2021. doi: 10.1103/PhysRevA.103.032430.

Yutaro Yamada, Ofir Lindenbaum, Sahand Negahban, and Yuval Kluger. Feature selection using stochastic gates. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10648–10659, 2020. URL https://proceedings.mlr.press/v119/yamada20a.html.

A Dataset definitions and preprocessing

Table 5: Canonical binary tasks and cached dimensions used in this study. The last column is the count of target-one samples; it does not always represent disease or an adverse outcome.

Dataset	Samples	Features	Target one
Hill-Valley	1212	100	606
Sonar	208	60	111
Heart	270	13	120
Diabetes	768	8	268
Ionosphere	351	34	225
Breast-W	699	9	241
PC4	1458	37	178
CNAE-9	1080	856	120
Breast Cancer	569	30	357
German Credit	1000	20	700
Blood Transfusion	748	4	178
Tic-Tac-Toe	958	9	626
QSAR	1055	41	356
Climate	540	20	494
Colon	62	2000	22
Leukemia	72	7129	25
GLI	85	22283	59
Ovarian	253	15154	91
Arcene	200	10000	112
DLBCL	77	5469	19

Table 5 summarizes the sample counts and feature dimensions of the twenty datasets. The cached feature and target arrays were hashed and matched across both execution hosts for all twenty datasets.

The task names refer to the canonical cached matrices, rather than every dataset with a similar repository name. Breast Cancer is the 569-sample diagnostic dataset distributed with scikit-learn, while Breast-W is the 699-sample, nine-feature task. CNAE-9 uses class "1" as the positive class and combines all other classes as negative. The high-dimensional tasks use the cached binary matrices with the dimensions in Table 5.

The fourteen conventional tasks are loaded from saved OpenML/scikit-learn arrays. Categorical labels are encoded consistently and nonnumeric input categories are numerically coded before CV. Missing or nonfinite entries remaining after canonical cleaning are replaced by zero. NCA and all data-dependent transformations in the LRS pipeline are fitted only on the corresponding training partition.

For LRS, a StandardScaler is fitted on the corresponding training partition. PCA uses a full singular-value decomposition and NCA is initialized from PCA, with at most 100 iterations and tolerance 10^{-5}. Each independently produces $k = \min(q, d, n_{\text{train}} - 1)$ coordinates, which are zero-padded to q before concatenation. A clipped MinMaxScaler is fitted to the concatenated training bank and mapped to $[-0.99\pi, 0.99\pi]$. Because padding precedes this map, padded zero columns become constant angles at -0.99π, not zero-angle inputs. They provide no sample information. The same fitted maps transform validation or test data without refitting.

NCA fit dimension and ordering differ between the full bank and count allocation. Learned selection is applied to fixed coordinates; it does not backpropagate through the PCA or NCA fitting algorithms.

B HYPERPARAMETERS AND TRAINING DETAILS

B.1 QUANTUM PIPELINES

Table 6 gives the search spaces actually used. A logarithmic range denotes log-uniform sampling by the TPE search. LRS and Count-12 use twenty trials per outer fold, with five initial random trials. LRS sampler seed is $101 + o + 200000$, where $o \in \{0, \ldots, 4\}$ is the zero-based outer-fold index. The model/preprocessing seed is $100o + i$ for inner fold i and $100o$ for outer refitting; here there is only inner fold $i = 0$. NCA uses that preprocessing seed plus one. The stratified inner holdout seed is $1101 + o$. Minibatch order uses a generator seeded with 42. These choices are reproducibility conventions, not independent seed replications.

Table 6: Hyperparameter search space and fixed settings.

Parameter	Range or value
Predictor learning rate	Logarithmic $[10^{-4}, 3 \times 10^{-2}]$
Encoding-scale learning rate	Logarithmic $[10^{-4}, 3 \times 10^{-2}]$
Source-logit learning rate (LRS)	Logarithmic $[10^{-4}, 3 \times 10^{-2}]$
Margin coefficient λ (LRS)	Logarithmic $[10^{-4}, 5 \times 10^{-2}]$
Predictor weight decay	Logarithmic $[10^{-8}, 10^{-2}]$
Source-logit weight decay	Zero
Positive-class weighting	None; balanced; clipped ratio
Count-12 allocation	$(p, n) = (0, 12), \ldots, (12, 0)$
Optimizer	Adam; separate predictor, scale, and source-logit groups
Batch size	16
Learning-rate schedule	Cosine annealing over the maximum stage updates
Stage cap	50 epochs for selection; 50 for fresh refit
Saturation stopping	Full training accuracy ≥ 0.99, checked every 5 epochs
Logit initialization	Uniform $[-0.05, 0.05]$
Surrogate temperature	$\tau_e = 2(0.3/2)^{(e-1)/49}$
Margin μ	1
Decision threshold	0.5
Angle mapping	Clipped $[-0.99\pi, 0.99\pi]$
Scale parameterization	$s_j = 0.1 + 2.9\,\mathrm{sigmoid}(r_j)$, initialized to 1
Circuit width/depth	12 wires; 4 rotation/ring layers
LRS wire groups	$(0, \ldots, 5)$ and $(6, \ldots, 11)$
Dynamic-group ablation	Reorder PCA-selected slots before NCA-selected slots; boundary at learned PCA count
Global-ring ablation	One ordered CNOT ring over all 12 wires
Width ablations	10 wires with 5+5 rings; 8 wires with 4+4 rings

For positive-class weighting, let $r = n_0/n_1$. The three options use $w = 1$, $w = r$, and $w = \min(2, \max(1, r))$, respectively. Loss probabilities are clipped to $[10^{-7}, 1 - 10^{-7}]$ before logarithms. The stage scheduler is indexed by minibatch updates, with its maximum based on the full stage cap even when saturation stopping occurs. The full training accuracy is computed in evaluation mode, using the hard gate choices.

Each inner trial performs both selection and fresh refitting before validation scoring. The best trial maximizes validation macro-F1; ties minimize binary cross entropy (BCE) and then trial number. The entire procedure, including feature fitting and source selection, is repeated on outer-training data using the selected hyperparameters. Both stages reset the predictor with the same model seed. Thus "fresh initialization" means reinitializing weights rather than continuing the selection-stage weights, not drawing a separately seeded initialization.

Count-12 uses the same holdout and twenty-trial budget but a single predictor-training stage. The first ring contains p wires when $0 < p < 12$; a single ring spans the register when $p = 0$ or $p = 12$. The second view is refitted with its requested number of output dimensions. Consequently, allocation also changes the circuit topology and NCA fit, unlike the fixed-bank LRS selector.

B.2 TREE SEARCH SPACES

Table 7 lists the search ranges for the three tree ensembles. The tree TPE sampler uses seed $101 + o$

Table 7: Fifty-trial tree search spaces, applied separately to raw and transformed representations. L denotes a logarithmic range; U denotes a uniform range.

Model/parameter	Candidates or range
All: class weight	None, balanced, clipped ratio as above
RF: trees	$100, 200, \ldots, 1200$
RF: maximum depth	Unlimited, $2, 4, 6, 8, 12, 16, 24, 32, 48$
RF: leaf/split minimum	Integers $\mathrm{L}[1, 32]$/$\mathrm{L}[2, 32]$
RF: criterion/bootstrap	Gini or entropy; true or false
RF: feature fraction	Square root, or $\mathrm{U}[0.05, 1]$
RF: bootstrap sample fraction	$\mathrm{U}[0.5, 1]$, conditional on bootstrap
Boosting: learning rate	$\mathrm{L}[0.005, 0.3]$
Boosting: L1/L2 regularization	Each $\mathrm{L}[10^{-8}, 100]$
Boosting: row/column fractions	$\mathrm{U}[0.5, 1]$/$\mathrm{U}[0.05, 1]$
XGBoost: maximum depth	Integers $[1, 12]$
XGBoost: minimum child weight	$\mathrm{L}[0.01, 64]$
XGBoost: minimum split loss	$\mathrm{L}[10^{-8}, 10]$
LightGBM: leaves	Integers $\mathrm{L}[2, 256]$
LightGBM: maximum depth	Unlimited, $2, 4, 6, 8, 12, 16$
LightGBM: minimum child samples	Integers $\mathrm{L}[1, 64]$
LightGBM: minimum child weight	$\mathrm{L}[10^{-5}, 10]$
LightGBM: minimum split gain	$\mathrm{L}[10^{-8}, 1]$

and ten startup trials. Hyperparameters are selected by mean inner three-fold macro-F1, with mean BCE and trial index as tie breakers. Boosting has a maximum of 2,000 rounds with patience 40, monitoring validation macro-F1. Final boosting rounds are the median selected count across inner folds. RF uses its selected tree count directly. Each estimator uses one CPU thread. Transformed-tree references use the archived PCA12/NCA12 bank fitted inside each inner or outer training partition, and therefore do not use the learned LRS mask.

C PER-DATASET RESULTS

All displayed results use split seed 101. A table entry is the unweighted mean over its completed outer test folds, followed by the sample standard deviation. Ranks are assigned independently for accuracy and macro-F1. Red, blue, and green indicate the first, second, and third distinct means after rounding to two decimals; tied entries receive the same color. The ranking applies only to the methods displayed in the relevant table and is not a significance test.

Tables 8 and 9 report dataset-level accuracy and macro-F1 for LRS, Count-12, and the original-feature tree ensembles.

Table 8: Accuracy (percent), fold mean ± sample standard deviation. Red, blue, and green indicate the first-, second-, and third-highest mean scores in each row, respectively. Scores tied after rounding to two decimal places share the same color.

Dataset	LRS	Count-12	RF-Raw	XGB-Raw	LGBM-Raw
Hill-Valley	**74.01±5.06**	**74.59±2.19**	57.35±2.40	55.20±1.30	**58.42±2.85**
Sonar	**84.12±5.78**	82.71±3.84	**85.59±3.72**	81.72±7.11	**85.57±3.39**
Heart	80.37±5.00	81.11±1.55	**82.96±3.31**	**83.33±1.85**	**82.22±5.00**
Diabetes	**76.31±2.74**	**75.66±2.92**	**75.65±2.70**	74.74±1.20	74.62±2.96
Ionosphere	**92.87±3.36**	92.60±4.21	**93.15±2.57**	**93.15±2.57**	**93.44±1.65**
Breast-W	**96.85±1.09**	**96.71±1.95**	**97.42±0.64**	**96.85±0.81**	**96.71±1.40**
PC4	**89.92±2.47**	87.24±3.00	88.96±2.36	**89.37±1.50**	**89.85±1.80**
CNAE-9	**99.17±0.39**	98.61±0.65	**99.07±0.33**	98.80±0.70	**98.98±0.39**
Breast Cancer	**96.66±0.74**	**96.48±1.39**	**97.01±2.31**	96.13±2.13	**96.48±2.57**
German Credit	68.80±2.14	70.60±3.47	**73.30±2.84**	**71.30±1.68**	**71.80±0.57**
Blood Transfusion	**76.09±8.54**	75.95±7.02	**76.35±5.14**	74.46±6.15	**77.02±5.08**
Tic-Tac-Toe	**98.33±1.19**	95.62±4.68	95.82±1.24	**99.58±0.68**	**100.00±0.00**
QSAR	**86.92±3.31**	**86.26±2.42**	**87.20±3.05**	86.16±2.91	**86.92±2.38**
Climate	**86.30±6.01**	**86.48±4.37**	**88.70±2.88**	**88.70±1.78**	**88.70±1.78**
Colon	**85.38±7.07**	**80.51±13.93**	**81.92±11.16**	**80.51±5.06**	78.72±12.94
Leukemia	**94.57±5.61**	92.00±11.93	**97.33±3.65**	**94.38±5.95**	93.14±6.91
GLI	81.18±9.67	82.35±9.30	**87.06±6.44**	**84.71±3.22**	**91.76±6.71**
Ovarian	**98.03±2.40**	**98.82±1.75**	97.63±2.14	96.45±3.77	**98.81±1.78**
Arcene	**81.00±7.20**	78.00±6.94	**84.00±2.85**	**84.50±3.26**	**81.00±5.76**
DLBCL	**96.08±3.58**	**93.58±7.66**	88.17±7.27	85.58±7.46	**89.50±3.88**
Mean (20)	**87.15**	86.29	**86.73**	85.78	**86.68**

Table 9: Macro-F1 (percent), fold mean ± sample standard deviation. Red, blue, and green indicate the first-, second-, and third-highest mean scores in each row, respectively. Scores tied after rounding to two decimal places share the same color.

Dataset	LRS	Count-12	RF-Raw	XGB-Raw	LGBM-Raw
Hill-Valley	**72.73±5.59**	**73.51±2.47**	57.04±2.75	55.15±1.26	**58.39±2.85**
Sonar	**83.87±6.07**	82.53±4.02	**85.27±4.06**	81.44±7.35	**85.39±3.46**
Heart	80.08±5.09	80.81±1.37	**82.60±3.20**	**82.88±2.00**	**81.91±4.95**
Diabetes	**74.16±3.80**	**73.53±2.98**	**73.68±3.07**	73.13±0.77	73.03±3.23
Ionosphere	92.14±3.84	91.65±4.95	**92.44±2.84**	**92.37±2.94**	**92.63±2.00**
Breast-W	**96.53±1.21**	96.35±2.18	**97.18±0.70**	**96.54±0.90**	96.42±1.49
PC4	76.63±5.26	72.65±3.05	**77.80±3.14**	**78.18±2.96**	**80.00±3.15**
CNAE-9	**97.83±1.06**	96.38±1.73	**97.63±0.86**	96.91±1.90	**97.43±0.99**
Breast Cancer	**96.41±0.80**	96.20±1.53	**96.80±2.47**	95.85±2.29	**96.22±2.75**
German Credit	62.11±3.57	63.46±5.24	**69.97±3.08**	**67.22±2.42**	**67.99±1.65**
Blood Transfusion	67.27±6.10	**67.70±4.62**	**67.90±4.51**	66.15±4.43	**68.06±4.50**
Tic-Tac-Toe	**98.14±1.33**	94.99±5.55	95.25±1.46	**99.54±0.76**	**100.00±0.00**
QSAR	**85.37±3.78**	84.74±2.65	**85.62±3.54**	84.43±3.35	**85.31±2.87**
Climate	67.50±8.49	59.77±4.13	**67.63±5.87**	**69.04±4.19**	**68.82±4.94**
Colon	**83.36±8.80**	**79.54±13.61**	**79.69±12.05**	77.97±6.01	76.66±13.77
Leukemia	**94.07±6.26**	91.64±12.18	**96.98±4.15**	**93.28±7.25**	92.04±8.18
GLI	73.55±19.22	**80.10±8.81**	**84.95±6.10**	79.33±6.31	**90.41±7.14**
Ovarian	**97.87±2.60**	**98.74±1.88**	97.45±2.26	96.23±3.91	**98.68±1.98**
Arcene	**80.58±7.24**	77.51±7.30	**83.55±3.24**	**84.11±3.67**	80.20±6.47
DLBCL	**95.00±4.60**	**92.39±8.55**	82.73±10.44	79.96±9.47	**84.88±6.31**
Mean (20)	**83.76**	82.71	**83.61**	82.48	**83.72**

C.1 Trees on the transformed bank

PN denotes the concatenated twelve-coordinate PCA and twelve-coordinate NCA bank. This uses more candidate coordinates than the selected twelve-coordinate LRS input and does not share its

selection mask. Tables 10 and 11 report the corresponding comparisons with trees trained on PCA12+NCA12 features.

Table 10: Accuracy (percent), fold mean ± sample standard deviation. Red, blue, and green indicate the first-, second-, and third-highest mean scores in each row, respectively. Scores tied after rounding to two decimal places share the same color.

Dataset	LRS	RF-PN	XGB-PN	LGBM-PN
Hill-Valley	74.01±5.06	**96.21±1.90**	**96.95±1.93**	**97.03±1.83**
Sonar	**84.12±5.78**	**81.74±2.64**	**82.66±6.29**	78.86±4.18
Heart	**80.37±5.00**	**80.00±4.42**	**77.78±4.72**	**80.00±4.42**
Diabetes	**76.31±2.74**	**75.14±4.16**	**75.65±3.26**	74.62±4.19
Ionosphere	**92.87±3.36**	**94.58±2.57**	**93.15±3.97**	**93.15±1.88**
Breast-W	**96.85±1.09**	**96.71±1.10**	**96.71±0.97**	**96.71±1.49**
PC4	**89.92±2.47**	**88.68±2.54**	88.06±2.43	**90.33±1.75**
CNAE-9	**99.17±0.39**	**99.17±0.60**	**98.89±0.70**	**98.70±0.69**
Breast Cancer	**96.66±0.74**	**96.49±1.64**	**95.95±2.38**	**95.95±2.55**
German Credit	**68.80±2.14**	**67.70±3.55**	**69.30±2.91**	**67.70±1.04**
Blood Transfusion	76.09±8.54	**76.35±4.49**	**78.62±3.84**	**77.02±7.32**
Tic-Tac-Toe	**98.33±1.19**	**98.23±1.08**	97.71±1.67	**97.91±1.74**
QSAR	**86.92±3.31**	**87.77±1.94**	**87.87±1.63**	**87.77±2.63**
Climate	**86.30±6.01**	**85.37±3.43**	82.59±3.90	**85.56±3.68**
Colon	**85.38±7.07**	**80.64±4.36**	**79.10±10.70**	76.03±7.07
Leukemia	**94.57±5.61**	**86.00±5.23**	**87.62±8.97**	79.24±8.40
GLI	**81.18±9.67**	**82.35±9.30**	77.65±13.41	**80.00±6.71**
Ovarian	**98.03±2.40**	94.49±3.51	**96.06±2.77**	**95.67±3.51**
Arcene	**81.00±7.20**	77.00±8.91	**79.50±5.97**	**78.50±3.79**
DLBCL	**96.08±3.58**	**92.08±5.61**	**88.50±9.25**	84.58±6.86
Mean (20)	**87.15**	**86.83**	**86.52**	85.77

Table 11: Macro-F1 (percent), fold mean ± sample standard deviation. Red, blue, and green indicate the first-, second-, and third-highest mean scores in each row, respectively. Scores tied after rounding to two decimal places share the same color.

Dataset	LRS	RF-PN	XGB-PN	LGBM-PN
Hill-Valley	72.73±5.59	**96.20±1.90**	**96.95±1.93**	**97.03±1.83**
Sonar	**83.87±6.07**	**81.57±2.72**	**82.57±6.34**	78.48±4.54
Heart	**80.08±5.09**	**79.65±4.25**	77.48±4.59	**79.75±4.35**
Diabetes	**74.16±3.80**	**72.93±3.33**	**73.45±3.31**	72.79±4.69
Ionosphere	92.14±3.84	**94.02±2.90**	**92.28±4.70**	**92.44±2.12**
Breast-W	**96.53±1.21**	**96.42±1.18**	**96.41±1.04**	96.39±1.63
PC4	**76.63±5.26**	**76.58±3.94**	75.07±4.67	**77.37±5.13**
CNAE-9	**97.83±1.06**	**97.80±1.69**	**97.03±1.98**	96.56±1.96
Breast Cancer	**96.41±0.80**	**96.20±1.78**	**95.64±2.58**	95.61±2.79
German Credit	62.11±3.57	**63.96±4.01**	**64.74±3.63**	**64.28±1.33**
Blood Transfusion	**67.27±6.10**	65.48±3.61	**68.68±4.93**	**66.54±5.42**
Tic-Tac-Toe	**98.14±1.33**	**98.03±1.21**	97.43±1.89	**97.64±1.98**
QSAR	85.37±3.78	**86.40±2.21**	**86.44±1.81**	**86.25±3.08**
Climate	**67.50±8.49**	**66.94±6.47**	64.01±5.91	**64.66±5.57**
Colon	**83.36±8.80**	**77.42±9.09**	**77.38±11.79**	73.07±7.64
Leukemia	**94.07±6.26**	**84.11±6.02**	**85.73±10.09**	76.99±10.05
GLI	**73.55±19.22**	**76.53±14.15**	70.63±18.06	**74.35±9.96**
Ovarian	**97.87±2.60**	93.98±3.98	**95.65±3.12**	**95.25±3.91**
Arcene	**80.58±7.24**	76.82±9.00	**79.16±5.89**	**78.23±3.81**
DLBCL	**95.00±4.60**	**89.49±7.06**	**84.81±11.07**	77.11±8.01
Mean (20)	**83.76**	**83.53**	**83.08**	82.04

C.2 Representation, topology, and width ablations

The following tables compare the main fixed-6+6 LRS with dynamic-group LRS, PCA-only 12, NCA-only 12, Ring-CNOT LRS, LRS-10, and LRS-8. Every model has 100 outer-test results on the same twenty datasets, and the ordered test-index hashes match for every fold. PCA-only and NCA-only use a single source block. Dynamic-group LRS orders the selected PCA coordinates before the selected NCA coordinates and sets the refitted circuit's group boundary to the PCA count. Ring-CNOT preserves slot order and uses one twelve-wire ring in the refitted predictor. LRS-10 and LRS-8 use fixed 5+5 and 4+4 wire groups, respectively. Dataset means and sample standard deviations use all five outer folds; the mean row weights all twenty datasets equally. Tables 12 and 13 give the complete dataset-level scores for these seven models.

Table 12: Accuracy (percent), mean ± sample standard deviation over five outer folds. All seven models use the same 20 datasets. Colors identify the top three distinct rounded means per row; the last row gives average dataset ranks.

Dataset	LRS	Dynamic	PCA-12	NCA-12	Ring	LRS-10	LRS-8
Hill-Valley	74.01±5.06	**74.59±1.16**	**75.83±2.20**	73.76±4.16	73.60±4.30	73.11±4.15	**75.50±3.22**
Sonar	**84.12±5.78**	83.61±6.79	84.10±6.38	82.22±2.68	**84.17±9.29**	**85.09±5.54**	83.15±6.39
Heart	80.37±5.00	79.63±5.56	**83.70±3.56**	81.11±5.46	**82.59±4.26**	**81.48±6.28**	80.74±7.12
Diabetes	**76.31±2.74**	74.48±1.98	74.87±3.83	**77.48±2.58**	74.48±3.00	75.00±3.13	**76.18±2.89**
Ionosphere	**92.87±3.36**	92.31±2.94	**94.58±3.84**	91.45±4.29	92.02±2.41	**93.45±2.38**	91.18±3.65
Breast-W	**96.85±1.09**	95.99±1.81	96.28±2.19	**97.14±0.89**	96.71±1.21	**96.99±1.29**	96.56±1.56
PC4	**89.92±2.47**	**88.82±2.19**	87.24±5.30	86.56±3.81	**89.92±1.76**	**88.06±4.89**	86.21±5.64
CNAE-9	**99.17±0.39**	**98.70±0.60**	93.43±4.01	98.61±0.65	**98.89±0.41**	98.61±0.57	79.17±38.28
Breast Cancer	96.66±0.74	**97.01±1.01**	**97.19±1.16**	96.66±1.15	95.96±2.37	**97.01±2.53**	**97.36±0.89**
German Credit	**68.80±2.14**	67.90±4.62	67.80±2.14	**70.10±2.77**	66.60±5.21	**70.20±1.35**	66.30±6.43
Blood Transfusion	**76.09±8.54**	75.55±8.80	**76.48±6.89**	75.68±8.11	**75.95±7.83**	**76.09±8.59**	75.41±8.37
Tic-Tac-Toe	**98.33±1.19**	97.81±2.88	97.39±3.23	97.39±1.23	**98.23±1.41**	**98.75±0.95**	83.40±5.68
QSAR	**86.92±3.31**	86.26±2.49	85.59±2.38	85.12±1.91	**87.30±1.88**	**86.64±1.48**	83.03±3.58
Climate	86.30±6.01	**86.48±6.60**	81.48±9.07	**88.15±5.98**	85.00±5.45	82.22±5.13	**87.41±3.38**
Colon	**85.38±7.07**	83.59±10.47	**83.97±5.25**	**87.05±4.54**	82.31±10.16	79.10±8.94	83.85±8.05
Leukemia	94.57±5.61	**97.24±3.79**	91.71±5.79	**97.33±3.65**	93.14±4.72	**94.67±7.30**	94.57±5.61
GLI	81.18±9.67	**88.24±4.16**	**89.41±7.67**	**88.24±4.16**	76.47±19.95	**85.88±3.22**	84.71±3.22
Ovarian	**98.03±2.40**	97.63±2.57	97.64±2.15	96.85±2.24	**98.03±1.96**	**98.42±1.64**	**98.04±1.96**
Arcene	**81.00±7.20**	76.50±6.27	**81.50±6.75**	**79.50±9.59**	77.50±10.00	73.50±14.32	74.50±5.97
DLBCL	**96.08±3.58**	83.58±29.44	**94.75±2.94**	**94.75±2.94**	**94.75±2.94**	93.50±0.23	**93.58±4.43**
Mean (20)	**87.15**	86.30	**86.75**	**87.26**	86.18	86.39	84.54
Average rank↓	**3.03**	4.55	**3.70**	3.90	4.25	**3.60**	4.98

Table 13: Macro-F1 (percent), mean ± sample standard deviation over five outer folds. All seven models use the same 20 datasets. Colors identify the top three distinct rounded means per row; the last row gives average dataset ranks.

Dataset	LRS	Dynamic	PCA-12	NCA-12	Ring	LRS-10	LRS-8
Hill-Valley	72.73±5.59	**73.46±1.52**	**74.83±2.39**	72.42±5.02	72.16±4.83	71.72±4.77	**74.42±3.54**
Sonar	**83.87±6.07**	83.39±6.97	83.66±6.75	82.10±2.65	**83.85±9.89**	**84.87±5.74**	82.97±6.57
Heart	80.08±5.09	79.49±5.57	**83.44±3.71**	80.91±5.37	**82.32±4.26**	**81.26±6.18**	80.53±7.10
Diabetes	**74.16±3.80**	72.38±2.96	72.96±3.80	**75.19±3.86**	71.38±3.48	72.67±2.78	**74.31±2.91**
Ionosphere	**92.14±3.84**	91.50±3.38	**93.94±4.49**	90.14±5.68	91.06±2.93	**92.70±2.73**	90.12±4.33
Breast-W	**96.53±1.21**	95.53±2.04	95.86±2.46	**96.86±0.98**	96.35±1.36	**96.68±1.44**	96.21±1.74
PC4	**76.63±5.26**	**73.80±4.11**	73.58±6.45	71.87±3.53	73.55±4.54	**75.19±5.27**	69.99±5.31
CNAE-9	**97.83±1.06**	**96.62±1.65**	82.34±12.87	96.31±1.95	**97.11±1.10**	96.26±1.60	69.15±39.36
Breast Cancer	96.41±0.80	96.79±1.10	**96.98±1.24**	96.40±1.27	95.60±2.65	**96.81±2.69**	**97.15±0.99**
German Credit	62.11±3.57	63.66±3.55	**63.90±1.70**	**64.60±2.29**	62.98±2.51	**65.22±3.35**	62.96±5.33
Blood Transfusion	67.27±6.10	66.33±6.23	**67.51±5.36**	64.83±4.14	**68.28±6.45**	67.08±6.12	**67.42±6.08**
Tic-Tac-Toe	**98.14±1.33**	97.57±3.23	97.16±3.47	97.11±1.32	**98.03±1.57**	**98.61±1.05**	81.42±6.82
QSAR	**85.37±3.78**	84.67±2.74	84.05±2.60	83.46±1.88	**85.84±2.07**	**85.22±1.86**	81.37±3.88
Climate	**67.50±8.49**	**68.26±8.19**	57.41±2.68	**68.37±4.56**	63.33±3.59	61.19±3.00	59.62±10.11
Colon	**83.36±8.80**	81.80±11.05	**82.61±5.58**	**86.13±4.67**	80.82±10.54	76.85±9.60	82.23±8.61
Leukemia	**94.07±6.26**	**96.92±4.23**	90.82±6.40	**96.98±4.15**	92.43±5.32	94.00±8.22	93.78±6.35
GLI	73.55±19.22	**85.78±4.64**	**87.39±9.13**	**85.78±4.64**	73.55±19.22	**81.75±5.30**	81.56±3.84
Ovarian	**97.87±2.60**	97.45±2.77	97.40±2.38	96.56±2.47	97.84±2.15	**98.28±1.80**	**97.88±2.12**
Arcene	**80.58±7.24**	76.00±6.20	**81.18±6.77**	**79.09±9.82**	77.19±9.83	73.03±14.40	74.35±5.96
DLBCL	**95.00±4.60**	82.30±30.25	**93.14±3.91**	**93.14±3.91**	**93.14±3.91**	91.31±0.79	**91.81±5.24**
Mean (20)	**83.76**	**83.18**	83.01	**83.91**	82.84	83.03	80.46
Average rank↓	**3.18**	4.43	**3.60**	3.98	4.23	**3.70**	4.90

Table 14 summarizes the paired comparisons between LRS and each ablation model.

Table 14: Paired dataset-level tests of LRS against the six ablation models. Δ is LRS minus comparator in percentage points; W/T/L counts wins, ties, and losses across 20 datasets. Holm correction covers all 12 tests.

Metric	Comparator	Mean Δ	W/T/L	Wilcoxon W	Holm p
Accuracy	Dynamic-group LRS	+0.85	15/0/5	48.0	0.329
Accuracy	PCA-only 12	+0.40	13/0/7	76.0	1.000
Accuracy	NCA-only 12	-0.11	12/0/8	95.0	1.000
Accuracy	Ring-CNOT LRS	+0.97	16/1/3	25.0	0.058
Accuracy	LRS-10	+0.76	9/1/10	77.0	1.000
Accuracy	LRS-8	+2.61	13/1/6	42.0	0.329
Macro-F1	Dynamic-group LRS	+0.57	14/0/6	67.0	1.000
Macro-F1	PCA-only 12	+0.75	12/0/8	84.0	1.000
Macro-F1	NCA-only 12	-0.15	12/0/8	89.0	1.000
Macro-F1	Ring-CNOT LRS	+0.92	15/1/4	38.0	0.240
Macro-F1	LRS-10	+0.73	11/0/9	84.0	1.000
Macro-F1	LRS-8	+3.30	12/0/8	56.0	0.539

The main model's mean accuracy and macro-F1 exceed Ring-CNOT by 0.97 and 0.92 percentage points, LRS-10 by 0.76 and 0.73 points, and LRS-8 by 2.61 and 3.30 points. Across the 12 paired tests, the smallest Holm-adjusted p-value is 0.058 for accuracy against Ring-CNOT. No comparison reaches the 0.05 threshold. NCA-only 12 retains slightly higher unweighted means, whereas the main model has the lowest average ranks across the seven models. Table 15 gives both metric ranks for every dataset.

Table 15: Per-dataset ranks across the seven ablation models. Each cell reports accuracy rank / macro-F1 rank; lower is better. Tied unrounded means receive average ranks.

Dataset	LRS	Dynamic	PCA-12	NCA-12	Ring	LRS-10	LRS-8
Hill-Valley	4 / 4	**3** / **3**	**1** / **1**	5 / 5	6 / 6	7 / 7	**2** / **2**
Sonar	**3** / **2**	5 / 5	4 / 4	7 / 7	**2** / **3**	**1** / **1**	6 / 6
Heart	6 / 6	7 / 7	**1** / **1**	4 / 4	**2** / **2**	**3** / **3**	5 / 5
Diabetes	**2** / **3**	6.5 / 6	5 / 4	**1** / **1**	6.5 / 7	4 / 5	**3** / **2**
Ionosphere	**3** / **3**	4 / 4	**1** / **1**	6 / 6	5 / 5	**2** / **2**	7 / 7
Breast-W	**3** / **3**	7 / 7	6 / 6	**1** / **1**	4 / 4	**2** / **2**	5 / 5
PC4	**1** / **1**	**3** / **3**	5 / 4	6 / 6	**2** / 5	4 / **2**	7 / 7
CNAE-9	**1** / **1**	**3** / **3**	6 / 6	4.5 / 4	**2** / **2**	4.5 / 5	7 / 7
Breast Cancer	5 / 5	4 / 4	**2** / **2**	6 / 6	7 / 7	**3** / **3**	**1** / **1**
German Credit	**3** / 7	4 / 4	5 / **3**	**2** / **2**	6 / 5	**1** / **1**	7 / 6
Blood Transfusion	**2.5** / 4	6 / 6	**1** / **2**	5 / 7	**4** / **1**	**2.5** / 5	7 / **3**
Tic-Tac-Toe	**2** / **2**	4 / 4	5 / 5	6 / 6	**3** / **3**	**1** / **1**	7 / 7
QSAR	**2** / **2**	4 / 4	5 / 5	6 / 6	**1** / **1**	**3** / **3**	7 / 7
Climate	4 / **3**	**3** / **2**	7 / 7	**1** / **1**	5 / 4	6 / 5	**2** / 6
Colon	**2** / **2**	5 / 5	**3** / **3**	**1** / **1**	6 / 6	7 / 7	4 / 4
Leukemia	4.5 / **3**	**2** / **2**	7 / 7	**1** / **1**	6 / 6	**3** / 4	4.5 / 5
GLI	6 / 6.5	**2.5** / **2.5**	**1** / **1**	**2.5** / **2.5**	7 / 6.5	**4** / **4**	5 / 5
Ovarian	**3.5** / **3**	6 / 5	5 / 6	7 / 7	**3.5** / 4	**1** / **1**	**2** / **2**
Arcene	**2** / **2**	5 / 5	**1** / **1**	**3** / **3**	4 / 4	7 / 7	6 / 6
DLBCL	**1** / **1**	7 / 7	**3** / **3**	**3** / **3**	**3** / **3**	6 / 6	**5** / **5**

C.3 Topology and width results by dataset

The grouped twelve-wire model exceeds Ring-CNOT in accuracy on 16 datasets, ties on one, and trails on three; the corresponding macro-F1 counts are 15/1/4. Its largest accuracy gains over Ring-CNOT occur on GLI (4.71 points), Arcene (3.50), and Colon (3.08). Ring-CNOT performs better on Heart (2.22 points), illustrating that cross-group connectivity can have dataset-specific effects.

Reducing the width to ten wires yields a smaller change in the cross-dataset means than reducing it to eight. LRS-10 achieves 85.09% accuracy on Sonar and 93.45% on Ionosphere, compared with 84.12% and 92.87% for the main model. The largest losses for LRS-8 occur on CNAE-9 (79.17% versus 99.17% accuracy; 69.15% versus 97.83% macro-F1) and Tic-Tac-Toe (83.40% versus 98.33% accuracy). The width comparison therefore reflects both a modest aggregate reduction at ten wires and stronger dataset-specific sensitivity at eight wires.

D Statistical comparisons

The Friedman omnibus p-values are 0.0297 for accuracy and 0.0134 for macro-F1. The common cohort contains 20 datasets. Table 16 reports the paired tests against the primary reference models.

Each paired observation is a dataset mean; the five folds are never treated as five independent datasets. Wilcoxon calculations round paired differences to ten decimal places, omit exact zero differences, and use the implementation's normal approximation when required by ties or zeros. Holm adjustment covers the eight primary-reference tests in Table 16 and, separately, the 12 ablation tests in Table 14. Bootstrap intervals use 20,000 resamples of paired dataset differences with random seed 101 and percentile endpoints. These unadjusted intervals address a mean-difference estimand, whereas signed-rank tests address a different paired-distribution property, so their conclusions need not coincide. Intervals and tests are exploratory because the tasks were not randomly sampled from a specified deployment population and some datasets have related domains.

Table 16: Exploratory paired comparisons of LRS minus each reference using complete dataset means. Differences and bootstrap intervals are in percentage points. Two-sided signed-rank tests use Holm correction jointly across all eight tests; W/T/L counts dataset wins/ties/losses.

Metric	Reference	Δ	95% interval	W/T/L	p_{Holm}
Acc.	Count-12	+0.85	[+0.15, +1.60]	14/0/6	0.607
Acc.	RF-Raw	+0.41	[-1.39, +2.66]	8/0/12	1.000
Acc.	XGB-Raw	+1.37	[-0.55, +3.82]	12/0/8	1.000
Acc.	LGBM-Raw	+0.46	[-1.53, +2.70]	9/2/9	1.000
F1	Count-12	+1.05	[-0.17, +2.24]	14/0/6	0.607
F1	RF-Raw	+0.15	[-2.18, +2.74]	7/0/13	1.000
F1	XGB-Raw	+1.27	[-0.91, +3.94]	11/0/9	1.000
F1	LGBM-Raw	+0.04	[-2.61, +2.62]	10/0/10	1.000

E SELECTION DIAGNOSTICS

Figure 4 shows the frequency of PCA selection at each input slot, and Table 17 summarizes the selected source counts by dataset.

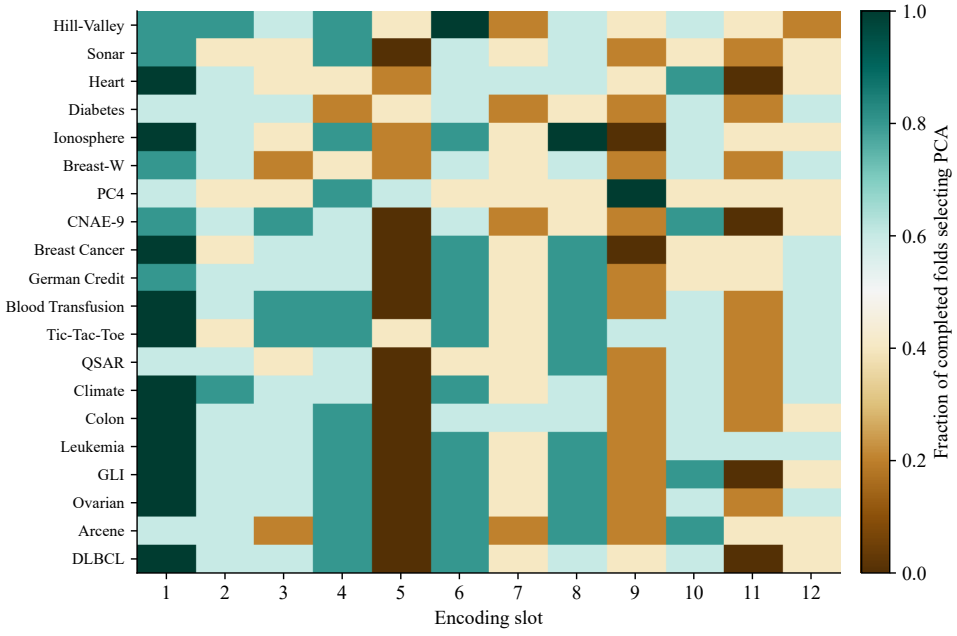


Figure 4: Fraction of completed outer folds selecting PCA at each slot. NCA occupies the complementary fraction. The mask is fixed within each fitted model. Shared seeds, coordinate order, and padded pairs influence the pattern; the heatmap is not causal feature importance or proof of selection stability.

Reconstructing the exact seeded initialization revealed 148 changed source choices among 1110 non-padded fold-slot pairs (13.3%); 52 of 100 final masks retained every initial choice on these slots. This diagnostic established that some choices changed, but also exposed substantial initialization retention. Table 18 reconstructs initial logits by instantiating the same architecture with each saved outer model seed, then compares their signs with saved final sources. Eligible slots omit padding. Identical masks at the endpoints do not exclude intermediate flips. Reporting the number of PCA choices alone does not characterize a representation, because masks with the same count can select different coordinates.

Table 17: Source selection across completed outer folds. PCA count is the mean number of PCA-selected slots; the mask is global within a fitted model. Low-dimensional padded slots can be uninformative.

Dataset	Folds	PCA count	Unique masks
Hill-Valley	5	6.80	5
Sonar	5	5.20	5
Heart	5	6.00	5
Diabetes	5	5.20	5
Ionosphere	5	6.60	5
Breast-W	5	5.40	5
PC4	5	6.20	5
CNAE-9	5	5.40	5
Breast Cancer	5	6.00	5
German Credit	5	6.20	5
Blood Transfusion	5	6.80	5
Tic-Tac-Toe	5	7.40	5
QSAR	5	5.40	5
Climate	5	6.40	5
Colon	5	6.20	5
Leukemia	5	7.00	5
GLI	5	6.40	5
Ovarian	5	6.60	5
Arcene	5	5.80	5
DLBCL	5	6.20	5

Table 18 compares the final source masks with their reconstructed initial values.

Table 18: Final versus reconstructed initialization masks. Padded slots are excluded. Changed choices count final sign differences, not the number of transitions during training; unchanged denotes folds with no final difference on eligible slots.

Dataset	Eligible slots	Changed choices	Unchanged folds
Hill-Valley	12	13/60	2/5
Sonar	12	7/60	3/5
Heart	12	13/60	2/5
Diabetes	8	15/40	0/5
Ionosphere	12	6/60	2/5
Breast-W	9	8/45	2/5
PC4	12	28/60	1/5
CNAE-9	12	10/60	2/5
Breast Cancer	12	5/60	4/5
German Credit	12	4/60	3/5
Blood Transfusion	4	1/20	4/5
Tic-Tac-Toe	9	8/45	1/5
QSAR	12	6/60	1/5
Climate	12	3/60	4/5
Colon	12	4/60	4/5
Leukemia	12	2/60	3/5
GLI	12	3/60	4/5
Ovarian	12	0/60	5/5
Arcene	12	8/60	2/5
DLBCL	12	4/60	3/5