Not All Quantum Gates Should Drop at Random: Fourier-Guided Regularization for Data-Reuploading QNNs

Pablo Rivas

Department of Computing Technology, School of Computer Science and Mathematics, Marist University, 3399 North Road, Poughkeepsie, NY 12601, USA
Pablo.Rivas@Marist.edu

Abstract. Data-reuploading quantum neural networks can express nonlinear classifiers through repeated input encoding, but this same capacity can overfit when only a few labeled examples are available. Standard quantum dropout masks circuit elements at random, ignoring the Fourier structure induced by the data-encoding gates. Here we study frequency-aware dropout, a training-time regularizer that scores reuploading blocks by their effect on learned spectral energy and masks blocks associated with unwanted frequency content. We test the method on exact-simulation binary classification tasks with low-, mixed-, and high-frequency decision structure, under clean labels, label noise, dropout-probability changes, qubit-count changes, and classical baselines. Soft frequency-aware dropout improves over random-gate and layer-block dropout in key mixed-frequency and noisy small-data settings, while preserving a clear failure mode on high-frequency targets. The results support Fourier-guided masking as an interpretable QNN regularization rule, not as a quantum-advantage claim.

Keywords: quantum machine learning · data re-uploading · quantum neural networks · quantum dropout · Fourier analysis · regularization

1 Introduction and Positioning

Quantum neural networks (QNNs) use parameterized quantum circuits as trainable models: a classical input is encoded by quantum gates, adjustable gate parameters are learned from labeled data, and a measurement is mapped to a prediction. A practical difficulty is that a circuit rich enough to fit a nonlinear boundary may fit spurious detail when the labeled set is small. Generalization under few training examples is therefore a central issue for QML models [1].

Data-reuploading QNNs make this issue both concrete and testable. They encode the same input repeatedly, alternating input-dependent blocks with trainable blocks. Pérez-Salinas et al. showed that repeated encoding can give a compact classifier a rich input response [7]. Such models are small enough for exact simulation, yet their repeated encodings provide enough capacity for regularization to matter.

A second observation gives this study its organizing idea. For common rotation encodings, a variational quantum model is a Fourier-type function of its input: the encoding design controls the available frequencies and trainable gates adjust their coefficients [9]. Thus, reuploading blocks are not merely interchangeable sets of gates. They contribute to the spectral structure of the learned decision function.

Dropout provides a natural regularization baseline. In QNN dropout, selected circuit operations are masked during training and the full circuit is evaluated at inference [8]. Recent work also considers softer quantum-dropout choices and quantum dropout outside supervised QNN classification [6,12]. These lines of work do not directly assign dropout probabilities from the Fourier behavior of data-reuploading blocks. We study that connection: *Fourier-guided dropout* scores each candidate reuploading block by how its temporary removal changes unwanted spectral energy, then masks blocks with score-dependent probabilities.

Relation to nearby work. The proposed rule is distinct from three established pieces. Data re-uploading supplies the circuit family; Fourier analyses explain its frequency response; and QNN dropout supplies stochastic regularization baselines. Repeated encoding has also been studied with qudits and in other variational tasks [11,3]. Our contribution is not any one of these components in isolation, but their use together: Fourier probes assign dropout probabilities to reuploading blocks during supervised training.

Our claim is bounded as follows: we do not claim that dropout is needed for every QNN, that a low-frequency prior is suitable for every target, or that these simulated QNNs outperform classical learning. We ask whether spectral information can make QNN dropout more useful and interpretable than random gate or random block masking in selected small-data settings.

We make three contributions. First, we define a block-level spectral-probe dropout rule for data-reuploading QNNs. Second, we test it against no dropout, random-gate dropout, and layer-block dropout on clean and label-noisy binary tasks with interpretable frequency roles. Third, we connect predictive accuracy with learned spectra and regularization metrics, while reporting limit cases through a high-frequency stress task, probability and qubit-count checks, and classical baselines. The supplementary appendix, placed after the references, provides full notation (Appendix B), implementation and protocol detail (Appendices C–D), additional results (Appendix E), and validity notes (Appendix F).

2 Fourier-Guided Dropout for Data-Reuploading QNNs

2.1 Base QNN and measured prediction

A qubit is a two-level quantum system whose state can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$. Gates are unitary transformations of this state. In a QNN, some gates carry data and others carry trainable angles. Measurement turns the final quantum state into a scalar used for classification.

For an input $x \in \mathbb{R}^d$, a depth-L data-reuploading circuit alternates an encoding block $S_\ell(x)$ and a trainable block $W_\ell(\theta_\ell)$,

$$U(x,\theta) = \prod_{\ell=1}^{L} W_\ell(\theta_\ell) S_\ell(x). \tag{1}$$

The encoding rotations reinsert the classical coordinates at every layer; the trainable rotations and, for multi-qubit cases, entangling operations transform the encoded state. The classifier measures Pauli-Z on the first wire,

$$z_\theta(x) = \langle 0|U^\dagger(x,\theta) Z_0 U(x,\theta)|0\rangle, \qquad p_\theta(y=1 \mid x) = \frac{1+z_\theta(x)}{2}. \tag{2}$$

Thus, an expectation value in $[-1,1]$ becomes a binary probability. All methods share Eq. (1), the loss, and the optimizer; they differ only in training-time masks.

2.2 Fourier structure and spectral bands

Rotation-based encoding makes the measured output periodic in its input angles. For the data encodings used here, the output can be expressed as a finite Fourier-type expansion [9],

$$f_\theta(x) = \sum_{\omega \in \Omega} c_\omega(\theta) e^{i\omega^\top x}. \tag{3}$$

Here, Ω denotes frequencies permitted by the encoding structure and $c_\omega(\theta)$ denotes learned coefficients. Repeated input encoding expands the frequency structure that the circuit can use. This is beneficial when a target needs richer variation, but under small or noisy data it can also fit sample-specific detail.

We compute Fourier coefficients from model outputs on a fixed evaluation grid and group them into low, mid, and high bands. For a band $\mathcal{B}$, its normalized energy is

$$E_{\mathcal{B}}(f_\theta) = \frac{\sum_{\omega \in \mathcal{B}} |c_\omega(\theta)|^2}{\sum_{\omega \in \Omega} |c_\omega(\theta)|^2}. \tag{4}$$

For one-dimensional models, low frequencies satisfy $|k| \le 2$, mid frequencies satisfy $3 \le |k| \le 5$, and higher frequencies form the high band. For two-dimensional models we use the analogous radial-frequency bands. For low- and mixed-frequency tasks, high-band energy is undesirable during scoring. The checker task deliberately applies this policy to a high-frequency target as a mismatch test.

2.3 Dropout rules and spectral probe

The comparison includes four modes (Fig. 1). No dropout trains the full circuit. Random-gate dropout masks eligible trainable rotations. Layer-block dropout masks complete $S_\ell(x) + W_\ell(\theta_\ell)$ units. Fourier-guided dropout masks the same full units, but its mask probabilities depend on spectral probes. In all dropout

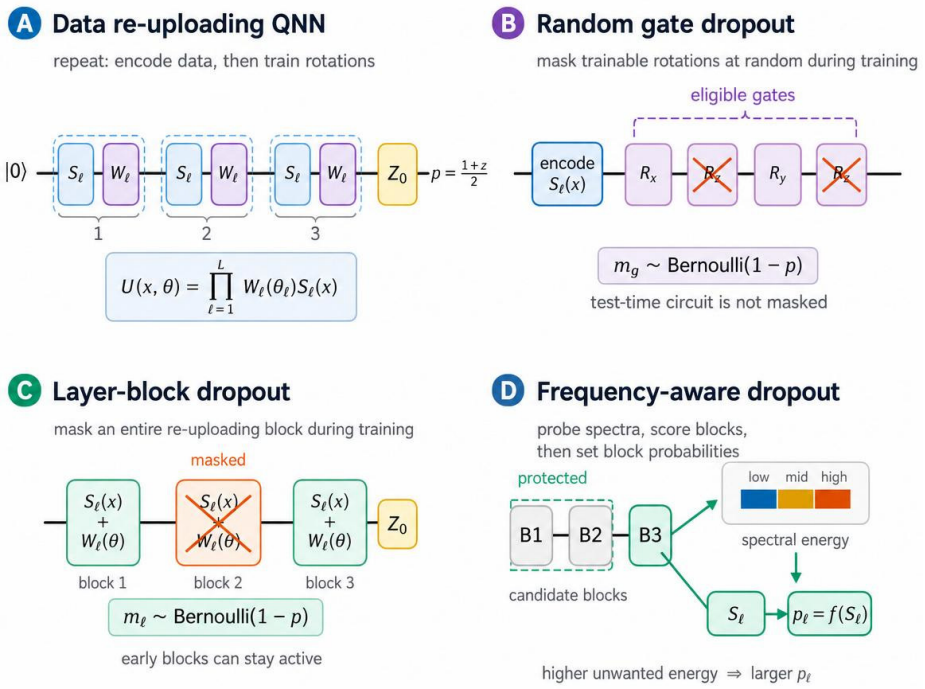

Fig. 1. The four training modes. All models share a data-reuploading circuit. Random-gate dropout masks trainable rotations; layer-block dropout masks complete reuploading blocks; Fourier-guided dropout assigns block probabilities using changes in unwanted spectral energy. Evaluation uses the full circuit.

modes, masks are sampled during training only; inference evaluates the unmasked circuit, following the training/evaluation distinction in dropout regularization [10,8].

Every scoring period, Fourier-guided dropout evaluates the current circuit on a fixed grid, temporarily omits each candidate block ℓ, and measures the reduction in undesirable band energy. Let $f_{\theta,-\ell}$ be the model with block ℓ omitted only for this probe. We define

$$s_\ell = \max\left(0, \frac{E_{\mathrm{bad}}(f_\theta) - E_{\mathrm{bad}}(f_{\theta,-\ell})}{E_{\mathrm{bad}}(f_\theta) + \epsilon}\right), \tag{5}$$

and convert this score into a mask probability,

$$p_\ell = \mathrm{clip}\left(p_{\min} + (p_{\max} - p_{\min})\frac{s_\ell}{\max_j s_j + \epsilon}, p_{\min}, p_{\max}\right). \tag{6}$$

Blocks whose temporary removal decreases unwanted energy are masked more often during subsequent optimization updates.

Table 1. Synthetic tasks and principal circuit settings. All synthetic test sets use 1024 examples and six reuploading layers. The checker task is the high-frequency stress case.

Dataset	Freq. role	d	Qubits	N_{train}	Main use
Low sine	low	1	1	32, 64	simple low-frequency control
Mixed sine	mixed	1	1	32, 64	mixed-frequency prior
Moons	mixed	2	2	32, 64	main 2D nonlinear task
Checker	high	2	2	32, 64	stress case

Fourier-guided training step. (1) Optimize the circuit with the current sampled block mask. (2) At a scheduled scoring epoch, evaluate the full circuit on the fixed grid and compute E_{bad}. (3) Omit each eligible reuploading block once, recompute E_{bad}, and obtain s_ℓ using Eq. (5). (4) Update p_ℓ using Eq. (6). (5) Continue training with block masks drawn from the updated probabilities. At test time, run the full unmasked circuit.

The clean core study showed that the first scoring range was too aggressive for some tasks. The final *soft* variant reduces the maximum mask probability and protects the earliest blocks, preserving initial input processing while still applying score-dependent regularization. Full masking settings, protected-layer behavior, grid definitions, and the complete training procedure are provided in Appendix C; the expanded Fourier notation is given in Appendix B.

3 Experimental Design

3.1 Tasks, circuits, and run families

We study four synthetic binary classification tasks chosen to test a frequency-based prior. For `sine_low_1d`, $y = \mathbb{1}[\sin(x) > 0]$. For `mixed_sine_1d`,

$$y = \mathbb{1}[\sin(x) + 0.5\sin(3x) > 0]. \tag{7}$$

The two-dimensional `moons_2d` task supplies a smooth nonlinear boundary after coordinate scaling to $[-\pi, \pi]$. The mismatch task, `checker_2d`, uses

$$y = \mathbb{1}[\text{sign}(\sin(3x_0)\sin(3x_1)) > 0]. \tag{8}$$

One-dimensional tasks use one qubit and two-dimensional tasks use two qubits in the main run; all use six reuploading layers. Each fixed split is shared across methods, $N \in \{32, 64\}$, and the synthetic test set contains 1024 points.

Table 2 records the empirical blocks. The clean experiment uses five seeds. Targeted follow-up studies use three seeds where stated in the supplementary protocol. The first Fourier-guided rule is retained as a diagnosis: it reduced generalization gaps but underfit some clean tasks. A pre-specified small soft-rule check selected the less aggressive setting used in the central figures.

Table 2. Completed experimental blocks retained in the evidence package.

Analysis block	Runs	Purpose
Clean QNN comparison	160	initial four-method test
Soft-rule check	54	correct over-regularization
Label noise	48	noisy-training robustness
Dropout probability	72	sensitivity to mask rate
Qubit count	24	two versus four qubits
Classical context	160	scope check

3.2 Optimization, metrics, and scope

All QNN results use exact statevector simulation, Adam optimization with learning rate 0.02, up to 150 epochs, full-batch updates, and validation-based early stopping with patience 25 [5]. No claim is made about finite-shot estimation or device noise. The noisy-label runs flip only training labels; validation and test labels remain clean.

Training minimizes binary cross-entropy,

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log p_\theta(x_i) + (1-y_i)\log(1-p_\theta(x_i))\right]. \tag{9}$$

We report accuracy as the primary predictive score and use train–test accuracy gap,

$$\Delta_{\text{gap}} = \text{Acc}_{\text{train}} - \text{Acc}_{\text{test}}, \tag{10}$$

plus Brier score and 10-bin expected calibration error (ECE) to assess probabilistic behavior [4]. Lower gap, Brier, and ECE are preferable when accuracy is comparable. Spectral band energies are computed from fixed-grid FFT evaluations of trained circuit outputs.

Classical logistic regression, RBF support-vector machines [2], and small MLP baselines are evaluated on the same splits. These models serve as context rather than as a target for an advantage claim. Full dataset construction, split rules, metric formulas, FFT grids, hyperparameters, and reproducibility records are given in Appendix D.

4 Results

4.1 Clean data: a soft rule is necessary

The initial exact-simulation study completed all 160 planned QNN runs. Its main lesson is that regularization strength matters. No dropout is a strong clean-data baseline, while the initial Fourier-guided policy reduces train–test gap but can suppress useful structure, most clearly on `moons_2d`. This diagnosis motivated a limited soft-rule check rather than an unrestricted method search: $p_{\max}$ was

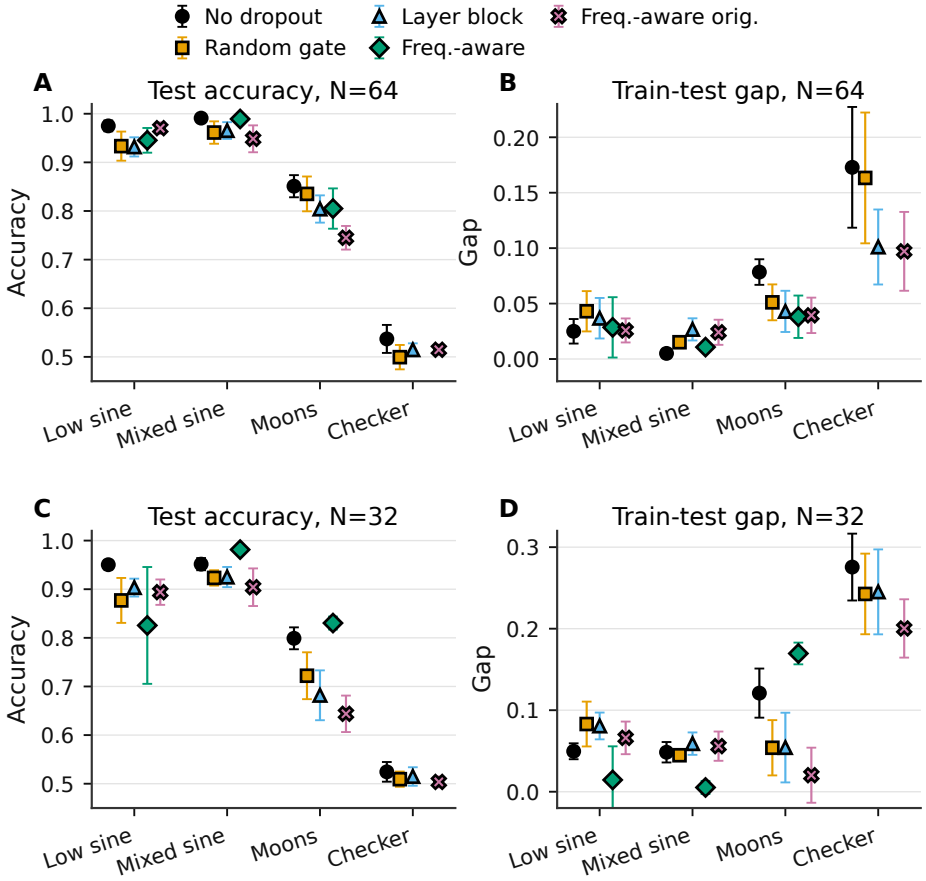

Fig. 2. Clean exact-simulation comparison and soft-rule refinement. No dropout is a strong clean-data reference; soft Fourier-guided dropout is most useful at small N on the mixed-frequency tasks and avoids the underfitting of the initial policy.

varied across three small values on the three non-stress tasks, for 54 completed runs.

The best overall soft-rule setting, $p_{\max} = 0.10$, improves the central dropout comparison. Figure 2 gives accuracy and gap across the clean settings, and Table 3 gives exact $N = 64$ accuracy values. On `mixed_sine_1d` at $N = 32$, soft Fourier-guided dropout reaches 0.981 ± 0.009, above no dropout (0.952 ± 0.013), random gate (0.923 ± 0.016), and layer block (0.925 ± 0.021). On `moons_2d` at $N = 32$, it reaches 0.830 ± 0.013, correcting the initial over-regularization and exceeding the other three methods. At $N = 64$, no dropout remains strongest on the three clean non-stress tasks, although Fourier-guided dropout nearly matches it on mixed sine and remains competitive with dropout baselines.

Table 3. Central clean-data prediction and calibration metrics at $N = 64$. Clean exact-simulation results for $N_{train} = 64$. Values are mean±SE across seeds. Bold marks the highest accuracy within each dataset. Lower is better for gap and ECE.

Dataset	Method	Acc.	Gap	ECE
Low sine	No dropout	**0.975±0.011**	0.025±0.011	0.080±0.016
Low sine	Random gate	0.933±0.030	0.043±0.018	0.121±0.012
Low sine	Layer block	0.932±0.020	0.037±0.018	0.143±0.011
Low sine	Freq.-aware	0.945±0.025	0.029±0.027	0.084±0.010
Mixed sine	No dropout	**0.991±0.003**	0.005±0.004	0.074±0.009
Mixed sine	Random gate	0.961±0.023	0.015±0.005	0.140±0.007
Mixed sine	Layer block	0.965±0.017	0.027±0.010	0.161±0.007
Mixed sine	Freq.-aware	0.989±0.001	0.011±0.001	0.085±0.019
Moons	No dropout	**0.851±0.023**	0.078±0.012	0.102±0.009
Moons	Random gate	0.835±0.036	0.051±0.016	0.095±0.016
Moons	Layer block	0.804±0.028	0.043±0.019	0.073±0.019
Moons	Freq.-aware	0.805±0.041	0.038±0.019	0.086±0.027
Checker	No dropout	**0.537±0.029**	0.173±0.054	0.158±0.031
Checker	Random gate	0.499±0.025	0.163±0.059	0.200±0.028
Checker	Layer block	0.515±0.013	0.101±0.034	0.169±0.013

The checker task gives an important boundary condition. All methods remain close to chance, with no dropout at 0.537 ± 0.029 and the original Fourier-guided rule at 0.515 ± 0.011 for $N = 64$. The scoring policy penalizes high-band content, while the checker boundary requires it; this is a planned mismatch result, not evidence that a low/mid-frequency prior should solve a high-frequency target.

4.2 Learned spectra support the design rationale

Figure 3 reports the learned spectral energy. The mechanism question is not whether lower high-band energy always yields higher accuracy; it does not. The question is whether structured masking changes spectral behavior in settings where the prior is relevant. On mixed sine and moons, the Fourier-guided rule limits excess high-band energy relative to standard dropout rules while retaining useful low/mid structure in the settings where its predictive results improve. On checker, the same bias does not provide an advantage, consistent with the task mismatch.

This result supports interpretation rather than a universal performance rule. A spectrum is a useful account of what the circuit learned, but accuracy also depends on optimization, sample variation, and the boundary required by the target. Full band-energy summaries and effective-degree results are reported in Appendix E, Tables E.3 and E.4.

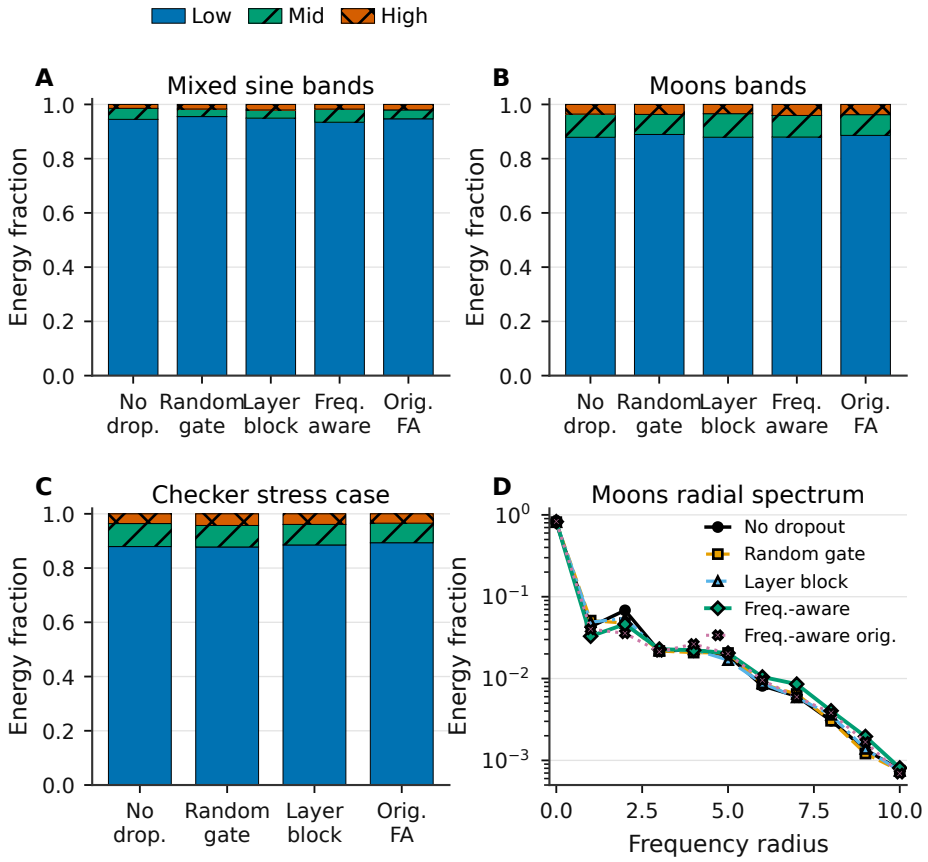

Fig. 3. Spectral evidence for the regularizer. Band-energy summaries show that Fourier-guided masking changes the learned spectrum in the mixed-frequency tasks; the checker stress task records the limit of applying a low/mid-frequency policy to a high-frequency target.

4.3 Calibration and gap analysis add a predictive-quality check

Accuracy alone does not say whether a small-data classifier is confident for the right reasons. We therefore measure both the train–test accuracy gap and probability quality through Brier score and expected calibration error (ECE). Figure 4 reports these measures for the clean comparison and the soft-rule settings. The principal point is not that one regularizer minimizes every metric; rather, Fourier-guided dropout improves predictive quality relative to the two random masking baselines in settings where its spectral prior is useful.

For `mixed_sine_1d` at $N = 32$, soft Fourier-guided dropout has Brier score 0.0228, compared with 0.0771 for random-gate dropout and 0.0726 for layer-block dropout; its ECE is 0.0835, compared with 0.1435 and 0.1315. At $N = 64$,

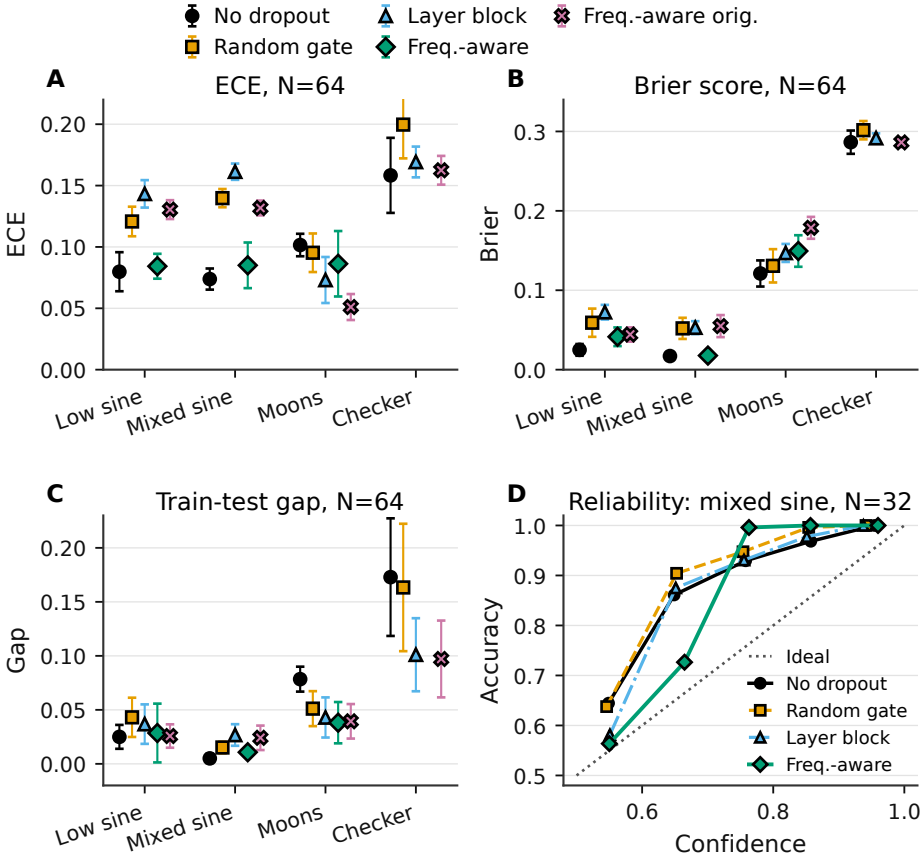

Fig. 4. Calibration and train–test gap checks. In central mixed-frequency comparisons, soft Fourier-guided dropout improves ECE or Brier score relative to random-gate and layer-block dropout, while the clean full circuit remains a strong reference. Full values are reported in Appendix E.

its Brier score is 0.0178, below random gate (0.0279) and layer block (0.0411). On `moons_2d`, the result depends on sample size: at $N = 64$, the soft rule reduces the train–test gap relative to no dropout (0.038 versus 0.069), whereas the $N = 32$ accuracy gain is accompanied by a larger gap. These observations rule out a one-metric success story and motivate reporting both predictive accuracy and regularization behavior. Complete calibration tables are provided in Appendix E, Tables E.1–E.2.

4.4 Training-label noise strengthens the structured-dropout case

The label-noise study completed 48 exact-simulation runs on mixed sine and moons. Figure 5 combines noise and selected ablation views; Table 4 records the

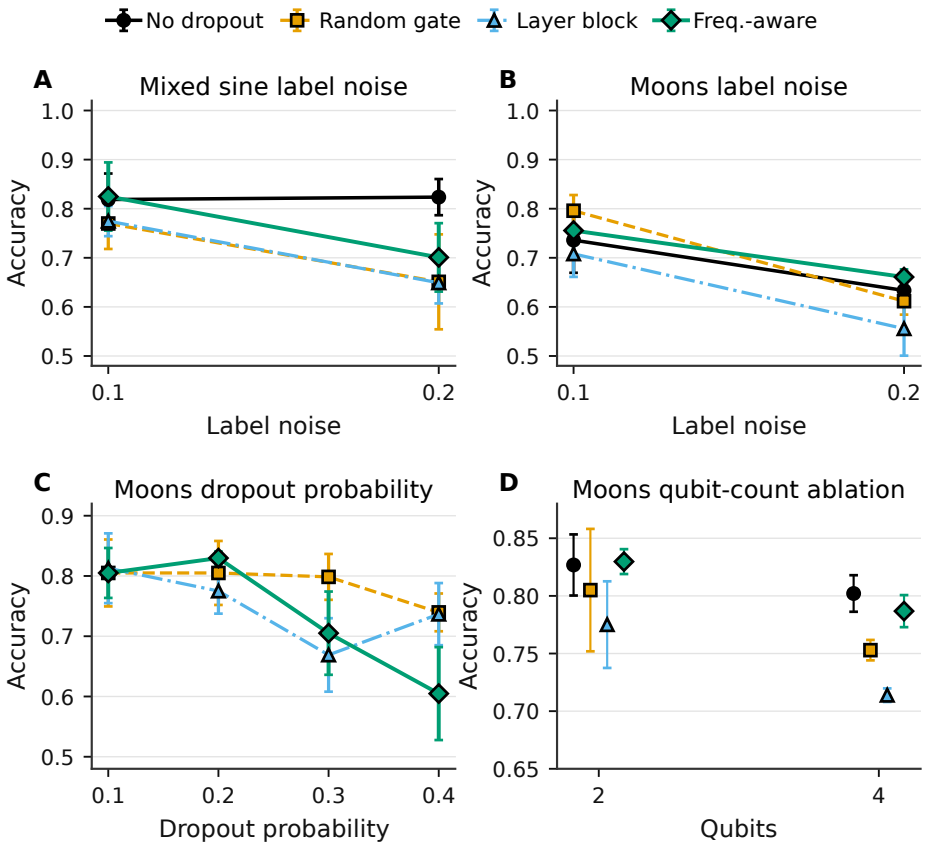

Fig. 5. Robustness and selected checks. Fourier-guided dropout gives its clearest gains when labels are noisy in mixed-frequency settings; the probability and qubit-count panels show that the useful regularization strength is task-dependent and that more qubits need not improve small-data performance.

central accuracy values. Fourier-guided dropout is the strongest dropout method on mixed sine under both noise rates and on moons at 20% noise. It also beats no dropout on mixed sine at 10% noise and on moons at 20% noise. For example, on noisy moons at 20%, it reaches 0.661 ± 0.016, compared with 0.633 ± 0.019 for no dropout, 0.612 ± 0.027 for random gate, and 0.556 ± 0.055 for layer block.

4.5 Checks that limit the claim

Three additional checks set the claim boundary. First, calibration analysis shows that soft Fourier-guided dropout can improve ECE or Brier score relative to random and layer-block dropout in key settings; detailed reliability displays and complete clean-data calibration values are given in Appendix E, Fig. 4 and Appendix Tables E.1–E.2. Second, the shared dropout-probability grid confirms a

Table 4. Prediction and spectral metrics with noisy training labels; test labels remain clean.

Dataset	Noise	Method	Acc.	ECE	High E
Mixed sine	10%	No dropout	0.819±0.053	0.065±0.013	0.044±0.015
		Random gate	0.770±0.052	0.092±0.041	0.057±0.011
		Layer block	0.775±0.031	0.074±0.026	0.059±0.005
		Freq.-aware	**0.825±0.069**	0.089±0.012	0.048±0.013
	20%	No dropout	**0.824±0.037**	0.109±0.035	0.059±0.017
		Random gate	0.651±0.097	0.140±0.074	0.064±0.012
		Layer block	0.649±0.042	0.152±0.048	0.075±0.020
		Freq.-aware	0.701±0.070	0.092±0.019	0.069±0.013
Moons	10%	No dropout	0.736±0.067	0.086±0.014	0.037±0.003
		Random gate	**0.796±0.032**	0.093±0.032	0.037±0.002
		Layer block	0.708±0.047	0.069±0.008	0.042±0.004
		Freq.-aware	0.756±0.003	0.055±0.005	0.043±0.002
	20%	No dropout	0.633±0.019	0.070±0.017	0.047±0.003
		Random gate	0.612±0.027	0.084±0.023	0.051±0.002
		Layer block	0.556±0.055	0.159±0.052	0.037±0.007
		Freq.-aware	**0.661±0.016**	0.064±0.018	0.040±0.006

Training-label noise results. Test labels are clean. Values are mean±SE across three seeds. Bold marks the highest accuracy in each dataset/noise group.

task-dependent effect: Fourier-guided dropout is the best dropout method on moons at its selected probability (0.830), whereas random-gate dropout is best on low sine (0.994). Third, on moons at $N = 64$, increasing the circuit from two to four qubits reduces accuracy for every QNN method; the Fourier-guided decrease is -0.043. These checks argue against any universal-win statement; their complete tabular values appear in Appendix E, Tables E.5 and E.6.

Classical baselines provide a final scope check. Table 5 records the $N = 64$ values needed to support our claim. On moons, an RBF SVM reaches 0.943, and on mixed sine a small MLP with dropout reaches 0.995. Thus, the result is not an advantage claim. It is a within-QNN regularization result: spectral information can select more useful dropout masks than random gate or layer-block dropout in selected small-data settings. Values at both training sizes are provided in Appendix E, Table E.7.

5 Discussion and Limitations

5.1 Interpretation

Data re-uploading gives compact QNNs a structured input dependence: repeated encodings make additional Fourier components available, while learned parameters determine their coefficients. The proposed rule uses that structure as a regularization signal. Rather than masking only isolated trainable gates at a

Table 5. Classical context on the $N = 64$ fixed splits. Values give context for QNN regularization; they do not support an advantage claim.

Task ($N = 64$)	Best classical model	Acc.	Brier
Low sine	Small MLP + dropout	0.989	0.008
Mixed sine	Small MLP + dropout	0.995	0.006
Moons	RBF SVM	0.943	0.044
Checker	Small MLP	0.516	0.256

fixed random rate, it estimates whether a whole reuploading block contributes unwanted spectral energy and adjusts that block's mask probability.

The evidence supports a constrained result. The first Fourier-guided policy was too strong on parts of the clean study; a softer policy corrected that failure. The soft rule is stronger than random-gate and layer-block dropout in key mixed-frequency and noisy-label settings, and the spectral analysis provides a direct account of the regularization behavior. The result is consistent with the broader concern that QML generalization under limited training data needs separate attention from trainability [1].

The negative results are central to that interpretation. No dropout is often strongest when the labels are clean. Random-gate dropout wins the simple low-frequency probability check. The checker task shows that a rule designed to discourage high-band energy is not suitable for a target that requires that content. Finally, the qubit-count check shows that more circuit width can worsen the test behavior of every compared method in a small-data setting.

5.2 Limitations

This study uses controlled synthetic classification problems. Their known frequency roles make them appropriate for testing a Fourier-guided rule, but they do not establish effectiveness on real inputs whose useful spectrum may be unclear after feature processing. The study is also based on exact simulation of small circuits. Finite-shot estimates would make both gradients and spectral probes noisy, and device noise could interact with masking in ways not tested here.

The soft policy was introduced after an initial policy underfit some tasks. We reduce the associated risk by reporting the initial behavior, limiting the repair to a small probability check, and retaining failure cases. Further validation should pre-register spectral bands or learn them only from a validation protocol. In addition, three-seed follow-up blocks are sufficient to identify large patterns but do not justify broad statistical claims.

The classical comparisons are also intentionally modest: standard models solve many of the synthetic tasks very well. Their role is to prevent an unsupported quantum-advantage reading. The valid contribution is narrower - a more structured dropout design inside the selected QNN family.

6 Conclusion

This paper studied dropout for data-reuploading quantum neural networks through the frequency structure induced by repeated data encoding. The central idea is simple: in a data-reuploading QNN, input-encoding gates help determine the frequencies that the model can express, so a dropout rule can use spectral information rather than masking circuit elements only at random.

The experiments show that this idea is useful, but not universal. The final soft frequency-aware dropout rule improves over random gate dropout and layer-block dropout in several mixed-frequency and noisy small-data settings. The spectral analyses support the proposed mechanism, and the calibration and label-noise results show that the benefit is not only an accuracy effect. The dropout-probability and qubit-count ablations also show that the strength of the regularizer matters: too much masking can underfit, and larger circuits can generalize worse under small data.

The limits are just as important as the positive results. No dropout remains a strong clean-data baseline. Random gate dropout can win on simple low-frequency tasks. Classical baselines are strong on these synthetic datasets. The high-frequency checkerboard task shows that a low/mid-frequency prior can be mismatched. The study is also simulation-only and does not test finite-shot hardware behavior.

The main conclusion is therefore a design rule, not a quantum advantage claim: when using compact data-reuploading QNNs on small datasets, dropout can be made more interpretable by tying masks to the spectral role of reuploading blocks. Future work should test richer spectral policies, finite-shot sampling, hardware noise, and higher-dimensional data where frequency bands are less direct but regularization remains critical.

Appendices B–F, placed after the references, report full notation, implementation and configuration detail, calibration and spectral tables, ablations, and reproducibility material.

Acknowledgments. Part of this work was funded by the National Science Foundation under grant CNS-2136961. The author thanks Dr. Gissella Bejarano Nicho for facilitating access to high-performance computing equipment. The author thanks the Rivas.AI Lab (`https://lab.rivas.ai`) for the support and helpful feedback throughout this project.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Caro, M.C., Huang, H.Y., Cerezo, M., Sharma, K., Sornborger, A., Cincio, L., Coles, P.J.: Generalization in quantum machine learning from few training data. Nature Communications **13**(1), 4919 (2022). `https://doi.org/10.1038/s41467-022-32550-3`
2. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3), 273–297 (1995). `https://doi.org/10.1007/BF00994018`

3. Cruz-Martinez, J., Robbiati, M., Carrazza, S.: Multi-variable integration with a variational quantum circuit. Quantum Science and Technology **9**(3), 035053 (2024). https://doi.org/10.1088/2058-9565/ad5866
4. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1321–1330. PMLR (2017). https://doi.org/10.48550/arXiv.1706.04599
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015), https://arxiv.org/abs/1412.6980
6. Lizzio Bosco, D., Romanello, R., Serra, G.: Softer is better: Tweaking quantum dropout to enhance quantum neural network trainability. In: 2025 IEEE International Conference on Quantum Computing and Engineering / Quantum Communications, Networking, and Computing. pp. 442–449. IEEE (2025). https://doi.org/10.1109/QCNC64685.2025.00076
7. Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I.: Data re-uploading for a universal quantum classifier. Quantum **4**, 226 (2020). https://doi.org/10.22331/q-2020-02-06-226
8. Scala, F., Ceschini, A., Panella, M., Gerace, D.: A general approach to dropout in quantum neural networks. Advanced Quantum Technologies (2023). https://doi.org/10.1002/qute.202300220
9. Schuld, M., Sweke, R., Meyer, J.J.: Effect of data encoding on the expressive power of variational quantum-machine-learning models. Physical Review A **103**(3), 032430 (2021). https://doi.org/10.1103/PhysRevA.103.032430
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**(56), 1929–1958 (2014), http://jmlr.org/papers/v15/srivastava14a.html
11. Wach, N.L., Rudolph, M.S., Jendrzejewski, F.: Data re-uploading with a single qudit. Quantum Machine Intelligence **5**(2) (2023). https://doi.org/10.1007/s42484-023-00125-0
12. Wang, Z., Zheng, P.L., Wu, B.: Quantum dropout: On and over the hardness of quantum approximate optimization algorithm. Physical Review Research **5**(2), 023171 (2023). https://doi.org/10.1103/PhysRevResearch.5.023171

Supplementary Appendix

The following appendices are provided part of the paper to add value to some sections.

A Supplementary Reading Guide

This supplementary appendix is written to help readers unfamiliar with quantum machine learning to enhance their understanding of the field. All this supplementary material studies a regularization question for small data-reuploading quantum neural networks (QNNs): when repeated data encoding gives a circuit a structured Fourier response, can that response be used to choose which circuit blocks are masked during training? The method considered here is *Fourier-guided dropout*, also referred to as frequency-aware dropout: at scheduled training epochs, each eligible reuploading block is temporarily removed, the change in unwanted spectral energy is measured, and the resulting score sets that block's future dropout probability. Evaluation always uses the full, unmasked circuit.

The appendix is organized as follows. Appendix B defines the QNN, data reuploading, Fourier representation, spectral bands, and effective spectral degree. Appendix C gives the four masking rules, the full scoring procedure, and the soft-rule setting used after the initial rule was found to be too aggressive. Appendix D records the datasets, splits, metrics, simulation settings, run families, and classical baselines. Appendix E reports the detailed calibration, spectral, probability, qubit-count, and classical-context evidence omitted from the short paper. Appendix F records validity limits and reproducibility information.

The appendix supports a bounded conclusion: Fourier-guided masking is a useful structured QNN regularizer in selected small-data settings where a low- or mixed-frequency prior is appropriate. It does not establish hardware performance, finite-shot robustness, superiority to classical models, or a rule that helps every target.

B Background and Full Notation

B.1 Qubits, gates, and measured classification

A qubit is a two-level quantum state,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1, \tag{B.1}$$

where the squared amplitudes give measurement probabilities. A circuit applies unitary gates to an initial state $|0\rangle^{\otimes n}$. In a variational QNN, data-dependent gate angles encode the input and trainable gate angles are optimized from examples.

For an input $x \in \mathbb{R}^d$, the depth-L circuit used in this study alternates a data-encoding block $S_\ell(x)$ and a trainable block $W_\ell(\theta_\ell)$:

$$U(x, \theta) = \prod_{\ell=1}^{L} W_\ell(\theta_\ell) S_\ell(x). \tag{B.2}$$

This repeated presentation of x is called *data re-uploading*; it allows compact circuits to express richer decision functions than a single input-encoding step [7].

The model reads a Pauli-Z expectation on wire 0 and maps it to the positive-class probability:

$$z_\theta(x) = \langle 0|^{\otimes n} U^\dagger(x,\theta) Z_0 U(x,\theta) |0\rangle^{\otimes n}, \qquad p_\theta(y=1 \mid x) = \frac{1+z_\theta(x)}{2}. \tag{B.3}$$

Since $z_\theta(x) \in [-1,1]$, Eq. (B.3) produces a probability in $[0,1]$. The predicted class is one when this probability is at least 0.5.

B.2 Exact circuit blocks used in the experiments

All principal QNN comparisons use six reuploading layers. For wire $w \in \{0,\dots,n-1\}$, let $a_w = 1 + 0.10w$. For a one-dimensional input $x = (x_0)$, each encoding block applies

$$R_X(a_w x_0) \longrightarrow R_Y(a_w x_0) \longrightarrow R_Z(0.05(\ell+1)a_w x_0)\,. \tag{B.4}$$

For two-dimensional inputs $x = (x_0, x_1)$, it applies

$$R_X(a_w x_0) \longrightarrow R_Y(a_w x_1) \longrightarrow R_Z(0.05(\ell+1)a_w (x_0+x_1))\,. \tag{B.5}$$

After each encoding block, every wire receives a three-angle trainable rotation $\mathrm{Rot}(\phi_{\ell w}, \vartheta_{\ell w}, \omega_{\ell w})$. For $n > 1$, neighboring wires are coupled by controlled-NOT gates, with a closing last-to-first connection when $n > 2$. The trainable tensor therefore has shape $L \times n \times 3$.

B.3 Why Fourier structure appears

A Pauli rotation includes trigonometric dependence on its angle. For example,

$$R_X(x) = \begin{bmatrix} \cos(x/2) & -i\sin(x/2) \\ -i\sin(x/2) & \cos(x/2) \end{bmatrix}. \tag{B.6}$$

When x enters such rotations and the circuit is measured, the output is a finite Fourier-type function. In one dimension this can be written as

$$f_\theta(x) = \sum_{\omega \in \Omega} c_\omega(\theta) e^{i\omega x}; \tag{B.7}$$

for vector inputs it becomes

$$f_\theta(x) = \sum_{\omega \in \Omega} c_\omega(\theta) e^{i\omega^\top x}. \tag{B.8}$$

The encoding generators and their repeated use determine the frequency support Ω, while trainable portions of the circuit control accessible coefficient values $c_\omega(\theta)$ [9]. This distinction supplies the reason to regularize reuploading blocks by spectral effect rather than by uniform random deletion alone.

B.4 Spectral summaries used in this study

For one-dimensional tasks, the frequency bands are

$$\mathcal{B}_{\text{low}} = \{k : |k| \le 2\}, \qquad \mathcal{B}_{\text{mid}} = \{k : 3 \le |k| \le 5\}, \qquad \mathcal{B}_{\text{high}} = \{k : |k| > 5\}. \tag{B.9}$$

For two-dimensional tasks, bands are defined by radial frequency $\|\omega\|_2$: low when $\|\omega\|_2 \le 2$, mid when $2 < \|\omega\|_2 \le 5$, and high otherwise. Given coefficients on the fixed FFT grid, normalized energy in band $\mathcal{B}$ is

$$E_{\mathcal{B}}(f_\theta) = \frac{\sum_{\omega \in \mathcal{B}} |c_\omega(\theta)|^2}{\sum_{\omega \in \Omega} |c_\omega(\theta)|^2}. \tag{B.10}$$

We also report an effective degree, the smallest radius containing at least 95% of total energy:

$$d_{95} = \min \left\{ d : \frac{\sum_{\|\omega\|_2 \le d} |c_\omega|^2}{\sum_{\omega \in \Omega} |c_\omega|^2} \ge 0.95 \right\}. \tag{B.11}$$

These summaries characterize learned functions; they are not assumed to determine accuracy by themselves.

C Full Model and Dropout Specification

C.1 Shared block notation and train/test behavior

Let $B_\ell(x, \theta_\ell) = W_\ell(\theta_\ell) S_\ell(x)$ denote the complete reuploading block in layer ℓ. All QNN methods use the same base circuit and optimizer. A block mask $m_\ell \in \{0, 1\}$ produces

$$U_m(x, \theta) = \prod_{\ell=1}^{L} B_\ell(x, \theta_\ell)^{m_\ell}, \qquad m_\ell \sim \text{Bernoulli}(1 - p_\ell). \tag{C.1}$$

Masks are active only during training. At validation and test time, every compared QNN is evaluated with the full circuit, $m_\ell = 1$ for all ℓ, consistent with dropout-style regularization [10,8].

C.2 Four compared QNN training modes

No dropout. No operation is masked; this model tests whether regularization is needed in each clean or noisy setting.

Random-gate dropout. Eligible single-qubit trainable rotations in W_ℓ are independently set to identity during training at a fixed probability. Data-encoding gates remain present. This is the gate-level stochastic baseline.

Layer-block dropout. Complete blocks B_ℓ are randomly omitted during training at a fixed probability, with protected early processing so that the model continues to receive encoded data. This baseline separates block masking from spectral scoring.

Fourier-guided dropout. Complete blocks are omitted as in layer-block dropout, but each eligible block obtains its own dropout probability from a spectral probe. For low- or mixed-frequency target policies, the undesirable energy is the high-band fraction $E_{\text{bad}} = E_{\text{high}}$. The checker experiment applies this same policy as an intentional mismatch case.

C.3 Spectral scoring rule

At each scoring epoch, let f_θ be the output of the full circuit on the fixed grid and let $f_{\theta,-\ell}$ be the output after block ℓ is temporarily omitted only for scoring. The nonnegative block score is

$$s_\ell = \max\left(0, \frac{E_{\text{bad}}(f_\theta) - E_{\text{bad}}(f_{\theta,-\ell})}{E_{\text{bad}}(f_\theta) + \epsilon}\right). \tag{C.2}$$

A block receives a larger score when removing it decreases the unwanted spectral fraction. Scores become dropout probabilities through

$$p_\ell = \text{clip}\left(p_{\min} + (p_{\max} - p_{\min})\frac{s_\ell}{\max_j s_j + \epsilon}, p_{\min}, p_{\max}\right). \tag{C.3}$$

Thus, the method does not remove high frequencies directly. It increases the probability of masking blocks whose temporary omission reduces the selected unwanted band.

C.4 Training algorithm

Algorithm A.1: Fourier-guided dropout training.

1. Initialize circuit parameters θ and block probabilities. Protect the required early layer or layers from block deletion.
2. For each training epoch, sample a block mask from current probabilities, evaluate binary cross-entropy on the training set, and update θ by Adam.
3. Every `score_period` epochs, evaluate the unmasked circuit on the fixed grid and compute $E_{\text{bad}}(f_\theta)$.
4. For each eligible block ℓ, temporarily omit only that block on the same grid, compute s_ℓ using Eq. (C.2), and restore the block.
5. Update the block probabilities using Eq. (C.3); continue stochastic training.
6. Select the saved model by the validation rule and evaluate train/test metrics with the complete unmasked circuit.

Table D.1. Dataset roles and circuit sizes. The stress task is intentionally mismatched with a policy that discourages high-band energy.

Dataset	Freq. role	d	Qubits	N_{train}	Main use
Low sine	low	1	1	32, 64	simple low-frequency control
Mixed sine	mixed	1	1	32, 64	mixed-frequency prior
Moons	mixed	2	2	32, 64	main 2D nonlinear task
Checker	high	2	2	32, 64	stress case

All synthetic test sets use 1024 examples and six reuploading layers. The checker task is the high-frequency stress case.

C.5 Soft-rule refinement and final policy

The initial Fourier-guided run used a stronger upper masking probability and revealed underfitting on portions of the clean comparison. A controlled follow-up reduced the maximum masking probability on the three non-stress tasks. The central soft setting is

$$p_{\max} = 0.10, \tag{C.4}$$

with early input-processing blocks protected. The later dropout-probability check reports task dependence rather than selecting a single universal value: for the moons check, the strongest Fourier-guided dropout result occurred at $p_{\max} = 0.20$, whereas on low sine the best dropout result came from random-gate dropout. The central claim therefore concerns a structured mask rule in suitable tasks, not a fixed probability that is best everywhere.

D Full Experimental Protocol and Hyperparameters

D.1 Synthetic tasks and fixed splits

The controlled tasks were selected because their spectral roles can be stated before training. In both one-dimensional tasks, x is sampled over $[-\pi, \pi]$. Low sine uses $y = \mathbb{I}[\sin(x) > 0]$, while mixed sine uses

$$y = \mathbb{I}[\sin(x) + 0.5\sin(3x) > 0]. \tag{D.1}$$

Moons is a noisy two-dimensional nonlinear classification task with features scaled to $[-\pi, \pi]$. Checker is a high-frequency stress case,

$$y = \mathbb{I}[\mathrm{sign}(\sin(3x_0)\sin(3x_1)) > 0], \tag{D.2}$$

for uniformly sampled $(x_0, x_1) \in [-\pi, \pi]^2$. Every method within a seed uses the same saved train, validation, and test split; class balance checks were applied during dataset creation.

Table D.2. Completed run families used in the study. Spectral and calibration records are computed from saved predictions or model evaluations rather than new comparative training grids.

Block	Scope	Seeds	Runs
Core exact QNN	4 tasks, $N = 32, 64$, 4 QNN methods	5	160
Soft-rule check	3 non-stress tasks, three soft settings	3	54
Spectral / calibration	Saved QNN models from the clean study	–	derived
Label-noise	Mixed sine and moons, $N = 64$, 10% and 20% training-label noise, 4 methods	3	48
Probability check	Low sine and moons, 3 dropout methods, four probabilities	3	72
Classical context	4 tasks, $N = 32, 64$, 4 classical models	5	160
Qubit-count check	Moons, $N = 64$, 2 and 4 qubits, 4 methods	3	24

D.2 Run families

Table D.2 lists each completed experimental block. The core comparison is the 160-run exact-simulation grid. The soft-rule check is a bounded method repair after the initial scoring range was found to suppress useful signal. The remaining blocks test mechanism, noisy labels, sensitivity, circuit width, and classical context.

D.3 Optimization and predictive metrics

For labels $y_i \in \{0, 1\}$ and model probabilities $p_\theta(x_i)$, training minimizes binary cross-entropy,

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log p_\theta(x_i) + (1 - y_i) \log(1 - p_\theta(x_i)) \right]. \tag{D.3}$$

We report test accuracy and train–test accuracy gap,

$$\Delta_{\text{gap}} = \text{Acc}_{\text{train}} - \text{Acc}_{\text{test}}\,. \tag{D.4}$$

Probability quality is assessed with the Brier score,

$$\text{Brier} = \frac{1}{N} \sum_{i=1}^{N} \left(p_\theta(x_i) - y_i \right)^2, \tag{D.5}$$

Table D.3. Principal optimization, simulation, and reporting settings.

Setting	Value
Simulator	Exact statevector simulation; no finite-shot or hardware-noise claim
Layers	Six reuploading layers in the main QNN studies
Qubits	One for 1D tasks; two for main 2D tasks; four only in the B2 check
Training sizes	$N_{\text{train}} \in \{32, 64\}$; fixed validation subset per seed
Test set	1024 clean synthetic examples per split
Optimizer	Adam, learning rate 0.02
Training limit	150 epochs; early stopping patience 25
Batching	Full-batch QNN optimization for these small training sets
FFT grids	256 equally spaced points for 1D; 64×64 grid for 2D
Noise study	Training-label flips only; rates 0.10 and 0.20
Central soft rule	Fourier-guided $p_{\max} = 0.10$ in central rescue/noise comparisons
Reported scores	Accuracy, train–test gap, Brier score, ECE, spectral energy, d_{95}

and expected calibration error with ten confidence bins I_b,

$$\mathrm{ECE}_{10} = \sum_{b=1}^{10} \frac{|I_b|}{N} \left|\mathrm{acc}(I_b) - \mathrm{conf}(I_b)\right|. \tag{D.6}$$

Lower gap, Brier, and ECE are preferable when predictive accuracy is comparable. Since ECE is bin-dependent, the reliability figure in Appendix E should be read together with Brier score and accuracy.

D.4 Spectral evaluation and noise protocol

After training, the circuit output is sampled on a fixed grid: 256 points for one-dimensional tasks and a 64×64 grid for two-dimensional tasks. FFT coefficients are grouped with the band rules in Eq. (B.9) and its radial two-dimensional counterpart, producing energy fractions and d_{95}.

In the robustness block, only training labels are flipped, independently at rates 0.10 and 0.20; test labels remain clean. Thus, any accuracy change is a response to noisy supervision rather than evaluation-label corruption.

D.5 Classical context and reproducibility record

Logistic regression, RBF support-vector machines [2], a small MLP, and a small MLP with dropout are evaluated on identical splits. They provide context only;

the study does not claim an advantage over classical learning. Each QNN experiment records configuration files, run status, prediction output, spectral summaries where relevant, and validation checks. The final archived results and generated tables are the source of record for the manuscript.

E Additional Results

This section supplies evidence that is useful for interpretation but not required to follow the short paper's central argument. All results remain within the same restricted claim: the soft Fourier-guided rule is useful relative to random gate and layer-block dropout in selected mixed-frequency or noisy-label settings, while clean/no-dropout and high-frequency limits remain visible.

E.1 Calibration and train–test gap

Tables E.1 and E.2 report mean accuracy, gap, Brier score, ECE, and high-band energy for the complete clean-data set used in this analysis. They supply the full numerical basis for the calibration view retained in the paper. The main observation is not that Fourier-guided dropout has the lowest calibration error everywhere. Rather, in key mixed-frequency comparisons it improves probability quality relative to the two random dropout baselines while remaining consistent with its predictive behavior.

Table E.1. Full clean-data calibration and gap summary for the one-dimensional tasks.

Dataset	Method	Acc.	Gap	Brier	ECE	High E
Low sine 32	No dropout	0.950	0.050	0.043	0.061	0.017
Low sine 32	Random gate	0.877	0.083	0.089	0.102	0.027
Low sine 32	Layer block	0.903	0.081	0.096	0.138	0.020
Low sine 32	Freq.-aware	0.826	0.014	0.128	0.104	0.029
Low sine 32	Freq.-aware original	0.894	0.066	0.096	0.110	0.023
Low sine 64	No dropout	0.975	0.025	0.025	0.080	0.016
Low sine 64	Random gate	0.933	0.043	0.059	0.121	0.022
Low sine 64	Layer block	0.932	0.037	0.072	0.143	0.018
Low sine 64	Freq.-aware	0.945	0.029	0.041	0.084	0.019
Low sine 64	Freq.-aware original	0.970	0.026	0.044	0.130	0.014
Mixed sine 32	No dropout	0.952	0.048	0.051	0.102	0.024
Mixed sine 32	Random gate	0.923	0.045	0.079	0.141	0.029
Mixed sine 32	Layer block	0.925	0.059	0.074	0.135	0.019
Mixed sine 32	Freq.-aware	0.981	0.005	0.023	0.084	0.012
Mixed sine 32	Freq.-aware original	0.904	0.056	0.083	0.112	0.026
Mixed sine 64	No dropout	0.991	0.005	0.017	0.074	0.015
Mixed sine 64	Random gate	0.961	0.015	0.052	0.140	0.017
Mixed sine 64	Layer block	0.965	0.027	0.053	0.161	0.021
Mixed sine 64	Freq.-aware	0.989	0.011	0.018	0.085	0.017
Mixed sine 64	Freq.-aware original	0.948	0.024	0.055	0.132	0.020

One-dimensional tasks. Mean values across seeds. Lower gap, Brier, ECE, and high-frequency energy are preferred only when accuracy remains comparable.

Table E.2. Full clean-data calibration and gap summary for the two-dimensional tasks.

Dataset	Method	Acc.	Gap	Brier	ECE	High E
Moons 32	No dropout	0.799	0.121	0.152	0.079	0.036
Moons 32	Random gate	0.722	0.054	0.190	0.073	0.046
Moons 32	Layer block	0.682	0.054	0.205	0.094	0.040
Moons 32	Freq.-aware	0.830	0.170	0.139	0.098	0.038
Moons 32	Freq.-aware original	0.644	0.020	0.225	0.077	0.044
Moons 64	No dropout	0.851	0.078	0.121	0.102	0.036
Moons 64	Random gate	0.835	0.051	0.131	0.095	0.037
Moons 64	Layer block	0.804	0.043	0.147	0.073	0.034
Moons 64	Freq.-aware	0.805	0.038	0.149	0.086	0.041
Moons 64	Freq.-aware original	0.745	0.039	0.179	0.051	0.038
Checker 32	No dropout	0.524	0.276	0.291	0.174	0.041
Checker 32	Random gate	0.509	0.243	0.297	0.185	0.040
Checker 32	Layer block	0.515	0.245	0.291	0.176	0.041
Checker 32	Freq.-aware original	0.504	0.200	0.304	0.194	0.042
Checker 64	No dropout	0.537	0.173	0.286	0.158	0.036
Checker 64	Random gate	0.499	0.163	0.302	0.200	0.043
Checker 64	Layer block	0.515	0.101	0.292	0.169	0.039
Checker 64	Freq.-aware original	0.515	0.097	0.286	0.162	0.035

Two-dimensional tasks. Mean values across seeds; checker is the high-frequency stress case.

E.2 Full spectral summary

Tables E.3 and E.4 expand the spectral evidence. For mixed sine at $N = 32$, soft Fourier-guided dropout combines the best displayed accuracy (0.981) with the smallest high-band fraction (0.012). On moons at $N = 32$, it improves accuracy to 0.830 without forcing the smallest high-band fraction. This distinction matters: spectral control supports the regularizer's interpretation, but it does not give a single monotonic predictor of test accuracy.

Table E.3. Full spectral summary for the one-dimensional tasks. Low, mid, and high entries are normalized FFT energy fractions; d_{95} is the smallest radius retaining at least 95% of spectral energy.

Dataset	Method	Acc.	Low E	Mid E	High E	d_{95}
Low sine 32	No dropout	0.950	0.929	0.054	0.017	3.200
Low sine 32	Random gate	0.877	0.919	0.054	0.027	3.400
Low sine 32	Layer block	0.903	0.939	0.041	0.020	3.000
Low sine 32	Freq.-aware	0.826	0.896	0.075	0.029	3.333
Low sine 32	Freq.-aware original	0.894	0.929	0.048	0.023	3.000
Low sine 64	No dropout	0.975	0.933	0.051	0.016	2.600
Low sine 64	Random gate	0.933	0.942	0.036	0.022	2.800
Low sine 64	Layer block	0.932	0.950	0.033	0.018	2.200
Low sine 64	Freq.-aware	0.945	0.914	0.067	0.019	3.000
Low sine 64	Freq.-aware original	0.970	0.940	0.047	0.014	2.600
Mixed sine 32	No dropout	0.952	0.940	0.036	0.024	3.400
Mixed sine 32	Random gate	0.923	0.940	0.031	0.029	3.600
Mixed sine 32	Layer block	0.925	0.942	0.039	0.019	2.600
Mixed sine 32	Freq.-aware	0.981	0.941	0.047	0.012	2.333
Mixed sine 32	Freq.-aware original	0.904	0.935	0.038	0.026	2.800
Mixed sine 64	No dropout	0.991	0.945	0.041	0.015	2.600
Mixed sine 64	Random gate	0.961	0.955	0.028	0.017	1.800
Mixed sine 64	Layer block	0.965	0.949	0.030	0.021	2.200
Mixed sine 64	Freq.-aware	0.989	0.934	0.048	0.017	3.000
Mixed sine 64	Freq.-aware original	0.948	0.947	0.033	0.020	2.600

One-dimensional tasks. Energy fractions sum to one up to rounding; standard errors are retained in frozen source data.

E.3 Dropout-probability and qubit-count checks

The probability check in Table E.5 limits a broad tuning claim. Fourier-guided dropout is the best dropout rule on moons at its best checked probability, but random-gate dropout is better on low sine. The qubit-count rows show that adding width from two to four qubits does not help the small-data moons task: accuracy declines for every compared QNN method.

Table E.4. Full spectral summary for the two-dimensional tasks. Low, mid, and high entries are normalized FFT energy fractions; d_{95} is the smallest radius retaining at least 95% of spectral energy.

Dataset	Method	Acc.	Low E	Mid E	High E	d_{95}
Moons 32	No dropout	0.799	0.893	0.071	0.036	4.344
Moons 32	Random gate	0.722	0.887	0.067	0.046	4.789
Moons 32	Layer block	0.682	0.889	0.071	0.040	4.310
Moons 32	Freq.-aware	0.830	0.902	0.061	0.038	4.095
Moons 32	Freq.-aware original	0.644	0.880	0.076	0.044	4.683
Moons 64	No dropout	0.851	0.879	0.086	0.036	4.346
Moons 64	Random gate	0.835	0.889	0.074	0.037	4.370
Moons 64	Layer block	0.804	0.879	0.087	0.034	4.237
Moons 64	Freq.-aware	0.805	0.880	0.080	0.041	4.572
Moons 64	Freq.-aware original	0.745	0.886	0.076	0.038	4.508
Checker 32	No dropout	0.524	0.869	0.089	0.041	4.683
Checker 32	Random gate	0.509	0.878	0.082	0.040	4.673
Checker 32	Layer block	0.515	0.878	0.080	0.041	4.530
Checker 32	Freq.-aware original	0.504	0.886	0.072	0.042	4.669
Checker 64	No dropout	0.537	0.879	0.085	0.036	4.532
Checker 64	Random gate	0.499	0.878	0.079	0.043	4.703
Checker 64	Layer block	0.515	0.885	0.076	0.039	4.683
Checker 64	Freq.-aware original	0.515	0.893	0.072	0.035	4.380

Two-dimensional tasks. Checker is an intentional frequency-policy mismatch test.

Table E.5. Dropout-probability and qubit-count summaries. These checks show task-dependent regularization and the absence of a simple more-qubits-is-better result.

A. Dropout-probability ablation.

Dataset	RG p	RG Acc.	LB p	LB Acc.	FA p	FA Acc.	FA-RG
Low sine	0.10	0.994	0.10	0.957	0.20	0.970	-0.024
Moons	0.10	0.805	0.10	0.813	0.20	0.830	0.025

B. Moons qubit-count ablation, 4 qubits minus 2 qubits.

Method	ΔAcc.	ΔECE	ΔGap
No dropout	-0.025±0.042	0.052±0.013	0.097±0.034
Random gate	-0.052±0.059	0.037±0.031	0.157±0.044
Layer block	-0.061±0.035	0.044±0.019	0.146±0.056
Freq.-aware	-0.043±0.022	0.042±0.001	0.069±0.059

RG: random gate; LB: layer block; FA: frequency-aware. Negative ΔAcc. means the four-qubit circuit was worse.

Table E.6. Detailed moons qubit-count result at $N_{\text{train}} = 64$. Values are mean±SE across three seeds; ranks are within each qubit count.

Qubits	Method	Acc.	Gap	ECE	Rank
2	Freq.-aware	0.830±0.011	0.059±0.037	0.104±0.003	1
2	No dropout	0.827±0.027	0.069±0.017	0.095±0.012	2
2	Random gate	0.805±0.053	0.025±0.005	0.088±0.025	3
2	Layer block	0.775±0.038	0.022±0.025	0.052±0.024	4
4	No dropout	0.802±0.016	0.165±0.017	0.147±0.003	1
4	Freq.-aware	0.787±0.014	0.128±0.028	0.146±0.002	2
4	Random gate	0.753±0.009	0.182±0.049	0.125±0.008	3
4	Layer block	0.714±0.006	0.168±0.037	0.096±0.005	4

Qubit-count ablation for Moons with $N_{train} = 64$. Values are mean±SE across three seeds.

E.4 Classical context

Table E.7 prevents a misreading of the result. Simple classical models are strong on these synthetic tasks: for example, the RBF SVM reaches 0.943 on moons at $N = 64$, and a small MLP with dropout reaches 0.995 on mixed sine at $N = 64$. The contribution is therefore a structured regularization result within the QNN family, not an advantage claim.

Table E.7. Best classical baseline per dataset and training size on the fixed splits. These scores are context for the QNN study, not a target claimed to be exceeded.

Dataset	N	Best classical	Acc.	ECE	Brier
Low sine	32	Small MLP + dropout	0.974	0.018	0.017
Low sine	64	Small MLP + dropout	0.989	0.012	0.008
Mixed sine	32	Small MLP + dropout	0.990	0.023	0.010
Mixed sine	64	Small MLP + dropout	0.995	0.015	0.006
Moons	32	RBF SVM	0.928	0.126	0.067
Moons	64	RBF SVM	0.943	0.060	0.044
Checker	32	Small MLP	0.507	0.077	0.260
Checker	64	Small MLP	0.516	0.068	0.256

Classical baselines provide context only. They are not used as a quantum-advantage target.

F Extended Validity and Reproducibility Notes

F.1 Scope of the evidence

The study concerns regularization within compact data-reuploading QNNs. It does not assert that Fourier-guided dropout is best for every task or that the QNNs are better than standard classical learners. The clean results, the checker stress case, the probability check, and the classical context table are included precisely because they limit the claim.

F.2 Controlled tasks and spectral policy

Synthetic tasks make a frequency-based question testable: their low-, mixed-, and high-frequency roles are specified before fitting. They also restrict external validity. In real data, a useful spectral prior may be unclear, especially after representation learning or dimensionality reduction. The checker result demonstrates the expected limitation of the policy used here: penalizing high-band energy is not suitable when that band is needed by the target.

The spectral metrics depend on fixed FFT grids and band thresholds. Different bands could alter energy fractions, so spectrum is reported as mechanism evidence rather than as a complete cause of accuracy. A future protocol could pre-register band rules for new benchmarks or select them using held-out validation only.

F.3 Simulation and model-family limits

All reported QNN training uses exact statevector simulation. This choice isolates masking from shot variance and device errors, but it leaves finite-shot spectral scoring and hardware noise untested. The circuit family is also small: one or two qubits in the central comparisons and four qubits only in one add-on check, with six reuploading layers. Results could change with other encodings, ansatz choices, depths, or connectivity patterns.

F.4 Method refinement and statistical limits

The soft Fourier-guided setting follows an initial policy that over-regularized some clean tasks. The manuscript makes that sequence explicit and reports the limited probability check rather than hiding the initial failure. The main exact grid uses five seeds, while follow-up blocks use three seeds. Means and standard errors summarize this controlled evidence, but the sample counts do not support universal or distribution-free statements.

F.5 Reproducibility record

The completed study preserves fixed dataset splits, per-run configuration files, run-status validation, predictions, spectral outputs, calibration summaries, figure-source files, and generated LaTeX tables. The paper displays were made from frozen summary inputs so that tables and figures do not depend on ad hoc recomputation. A full source archive should accompany any final release so readers can inspect mask probabilities, spectral scoring, and the exact evaluation configuration.

F.6 Summary

The appropriate reading of the combined evidence is narrow and positive: in controlled small-data settings with a matching low- or mixed-frequency prior, Fourier-guided dropout provides an interpretable block-masking rule that can be stronger than fixed random QNN dropout. Tests under finite shots, hardware noise, larger circuits, and less controlled data remain future work.