EVIDENTIAL UNCERTAINTY QUANTIFICATION WITH QUANTUM NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce a Quantum Evidential Generative Network (QEGN) that assigns quantum neural networks (QNNs) a complementary role as uncertainty quantification heads for classical classifiers. Belief function theory provides a shared semantics: the classical network supplies predictive tendencies expressed as singleton contours, while the quantum network learns higher order relations through evidence assigned to subsets of hypotheses. The proposed method finds a novel role for QNN in machine intelligence, which compacts relational guidance supports learning without a complete target mass distribution. We further extend evidential semantics to general quantum circuits through evidence correction with unitary evolution and give gates an explicit evidential interpretation. Experiments on relation learning, multimodal fusion, and uncertainty quantification support the practical use of the generated evidence.

1 INTRODUCTION

Quantum deep learning has become a promising direction in quantum machine intelligence, using trainable circuits and quantum-state representations for learning (Biamonte et al., 2017; Cerezo et al., 2021). Classical learning ideas have been translated into quantum models by encoding features in quantum states for kernel classification (Havlíček et al., 2019), implementing convolution and pooling through quantum circuits (Cong et al., 2019), and modeling data distributions with Born probabilities (Liu & Wang, 2018). However, a key design question remains: **how should classical and quantum models be coordinated in practical applications, and what complementary roles should they play in representing and processing information?**

There are two prerequisites from an information-processing perspective. At the representation level, classical information must be accurately encoded into quantum states without preparation costs offsetting the benefits of hybrid computation. At the processing level, classical updates and quantum evolution need a common semantics explaining how the represented information is revised. Quantum Bayesian approaches connect the two representations via Born probabilities (Low et al., 2014; Borujeni et al., 2021), which encode conditional probabilities through controlled rotations and implement inference through amplitude amplification. **However, preparing an unstructured distribution can require gate resources exponential in the number of qubitse, and updating process does not define a general informational meaning of quantum gates.**

Fortunately, there is an alternative route, belief function theory, to these prerequisites by assigning belief mass to subsets of hypotheses (Shafer, 1976). Figure 1 illustrates the correspondence between this power-set representation and qubit-based computation: qubits encode hypothesis membership, computational-basis strings identify subsets, and Born probabilities recover their belief masses (Zhou et al., 2023). Under this correspondence, logical gates implement set operations with explicit evidential meaning, and selected fusion and correction subroutines have linear gate counts for prepared inputs (Zhou et al., 2026). **Therefore, we adopt belief function theory as a unified semantic framework for classical-quantum information processing to develop corresponding machine intelligence methods.**

Belief function theory is an useful tool for epistemic uncertainty representation. Denœux (2019) showed that a neural network's softmax output layer can be interpreted as combining simple mass functions, exposing distinctions between missing and conflicting evidence beyond its class probabili-

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.

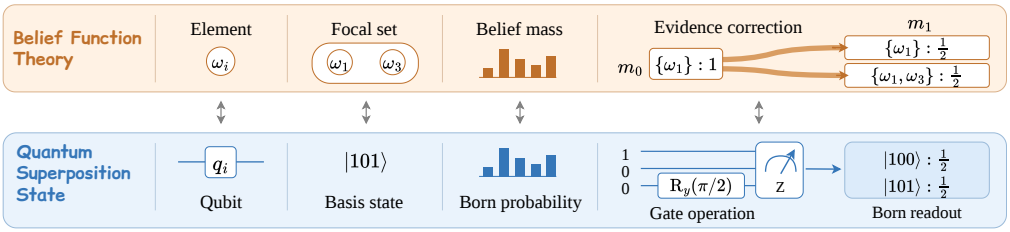

Figure 1: Belief functions connect classical and quantum information processing. Each qubit records membership of an elementary hypothesis, basis strings index focal sets, and computational-basis Born probabilities recover belief masses.

ties. Evidential deep learning (EDL) represents class-specific evidence through a Dirichlet distribution and summarizes uncommitted belief as global ignorance (Sensoy et al., 2018). Its standard formulation, however, cannot independently express ignorance confined to arbitrary subsets of output classes. Hyper-Evidential Neural Networks and Random-Set Neural Networks extend this representation through composite classes and selected focal sets (Li et al., 2024a; Kudukkil Manchingal et al., 2025). Assigning a separate node to every subset would make the output dimension exponential in the number of classes, limiting the practical use of richer evidential representations. **Therefore, classical evidential models are constrained by the difficulty of adequately representing high-dimensional information.**

We propose a *Quantum Evidential Generative Network* (QEGN) that combines quantum state representations with belief function semantics for epistemic uncertainty. As a quantum uncertainty head for a classical classifier, QEGN connects efficient information encoding, complementary representation learning, and interpretable circuit evolution. Our contributions are threefold:

1. We revisit the relationship between classical and quantum neural networks by introducing QNNs as complementary uncertainty quantification heads for classical predictors.
2. We establish a shared belief function semantics in which the classical network learns contours describing predictive tendencies, while the quantum network learns higher order relations among output hypotheses.
3. We extend evidential semantics to general quantum circuits by interpreting arbitrary unitary gates and their compositions as evidence correction operations.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 develops the quantum–classical uncertain-information framework, and Section 4 presents QEGN and its circuit-learning procedure. Section 5 reports the experiments, followed by discussion and conclusions in Section 6. The appendices provide further derivations, a three-qubit example, and implementation details.

2 Related Work

Quantum machine intelligence and complementary roles. Quantum learning now spans kernel methods, attention networks, ensembles, compression, and generative modeling (Cerezo et al., 2022; Fan et al., 2026). Building on quantum feature-space classification (Havlíček et al., 2019), Xu et al. (2024) learn feature mappings that cluster samples by class and supply trainable kernels for support-vector classification. Shi and collaborators construct self-attention from quantum query, key, and value states in QKSAN (Zhao et al., 2024); their recent QTrans preprint integrates circuit-generated attention features and a quantum feed-forward network into a hybrid transformer (Zhao et al., 2026). Dong and collaborators study mixed-reference quantum autoencoders for state compression (Ma et al., 2024) and establish a binary prediction-risk bound for quantum AdaBoost ensembles of variational quantum classifiers (Wang et al., 2024). Generative models have also progressed beyond the original circuit Born machine (Liu & Wang, 2018): nonlinear quantum neurons introduce measurement-conditioned activations (Gili et al., 2023), while quantum generative diffusion learns mixed-state targets through quantum channels and trainable reverse evolution (Chen et al., 2026b). Current

training and circuit-design workflows often use classical computation to compensate for practical limitations of quantum processing (Jerbi et al., 2023; Chinzei et al., 2025). Such coordination primarily supports quantum training and execution, while leaving the informational roles of classical and quantum models less explicitly defined.

Belief functions and uncertainty quantification. Belief functions represented incomplete knowledge by assigning mass to sets of hypotheses, allowing uncertainty about a subset without committing to its individual members (Shafer, 1976; Smets & Kennes, 1994). Denœux (2019) interpreted logistic regression and neural softmax classifiers as combinations of simple evidence, revealing missing and conflicting support behind the same probability output. EDL turned neural evidence into a Dirichlet distribution with an explicit global ignorance component (Sensoy et al., 2018). Recent improvements reweighted evidence-learning losses using Fisher information (Deng et al., 2023), relaxed fixed prior weights and variance-minimization regularization in R-EDL (Chen et al., 2024), incorporated feature-space density in DAEDL (Yoon & Kim, 2024), and suppressed spurious evidence with information-bottleneck regularization for fine-tuned language models (Li et al., 2025). Surveys placed these approaches within a wider landscape of evidential and random-set uncertainty representations (Gao et al., 2026). Zhang et al. (2023) learned composite image classes to express partial ignorance, while Li et al. (2024b) learned singleton and compound class masses from precise and ambiguous samples with an evidential radial-basis-function network. HENN represented partitioned composite classes through grouped Dirichlet distributions (Li et al., 2024a), and RS-NN learned belief values over a selected family of focal sets (Kudukkil Manchingal et al., 2025). Other studies used this distinction operationally: Zhao et al. (2025) combined ignorance and conflict estimates for active sample acquisition, and Su et al. (2025) separated non-specificity and conflict in an entropy measure for cooperative fusion. Huang (2024) combined complementary information from labeled and unlabeled data; Huang et al. (2025) learned class-dependent modality reliability through contextual discounting for multimodal segmentation. These studies illustrated how belief functions retained unresolved alternatives and qualified evidence before reasoning.

Quantum computation and evidential reasoning. Quantum-inspired evidential frameworks introduce complex-valued masses and phase-dependent fusion (Xiao et al., 2024). Chen et al. (2026c) combines complex-valued contrastive learning with generalized quantum evidence for multi-view inference. They do not establish a computational advantage from executing quantum circuits. The quantum transferable belief model extends this correspondence to fusion, correction, product-space operations, and evidential reasoning (Zhou et al., 2026). For prepared evidence states, selected fusion and correction operations use compact element gate modules. Xiao et al. (2025) proposed AQC-DRC implements Dempster's combination rule through adaptive Toffoli-based circuits. Applications include attribute-level evidential classification (Luo et al., 2026a), multimodal feature fusion through reliability-weighted conjunction (Wu et al., 2026) and vertical federated learning (Luo et al., 2026b). Together, these constructions show the potential of shared belief function semantics to connect encodings, but their interpretations attach to designed evidence operations and fusion modules; a compositional account of general learned circuit evolution remains a further step.

3 A QUANTUM–CLASSICAL FRAMEWORK FOR UNCERTAIN INFORMATION PROCESSING

Belief functions connect classical uncertain information with quantum states through subset membership and Born probabilities. We state the unitary evidence-correction representation and its properties below; their derivations and proofs are given in Appendix A.

3.1 QUANTUM TBM: ENCODING AND UPDATING UNCERTAIN INFORMATION

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a frame of hypotheses. A basic probability assignment (BPA) satisfies $m(A) \geq 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. Subsets with positive mass are focal sets: composite ones support unresolved alternatives, while $m(\varnothing)$ records conflict in the TBM (Smets & Kennes, 1994). Assigning one qubit to each hypothesis's membership, the basis state $|A\rangle = |z_1(A) \cdots z_n(A)\rangle$, with $z_i(A) = \mathbf{1}\{\omega_i \in A\}$, indexes a subset. The canonical mass function quantum state (MFQS) is (Zhou

et al., 2023; 2026)

$$|m\rangle = \sum_{A\subseteq\Omega} \sqrt{m(A)}|A\rangle, \qquad m(A) = |\langle A \mid m\rangle|^2. \tag{1}$$

Quantum TBM implements prescribed evidence rules through transformations of subset registers (Zhou et al., 2026). For conjunctive fusion, a unitary retains both source labels and computes their intersection in an output register initially labeled $\varnothing$:

$$U_\cap|A\rangle|B\rangle|\varnothing\rangle = |A\rangle|B\rangle|A\cap B\rangle. \tag{2}$$

For independent MFQS inputs, let $|\Psi_\cap\rangle = U_\cap(|m_1\rangle \otimes |m_2\rangle \otimes |\varnothing\rangle)$. Measuring its output register gives

$$m_\cap(F) = \langle\Psi_\cap|(I_{\text{src}} \otimes |F\rangle\langle F|)|\Psi_\cap\rangle = \sum_{A\cap B=F} m_1(A)m_2(B). \tag{3}$$

Here I_{src} acts on the retained source registers, and the outcome $F = \varnothing$ preserves conflict. Singleton contours and commonalities are membership observables: $\pi_i(m) = \langle(I - Z_i)/2\rangle$ and $q_m(S) = \langle\prod_{i:\omega_i\in S}(I - Z_i)/2\rangle$, where Z_i is the Pauli-Z operator on qubit i. Quantum TBM thereby connects mass encodings with rule-defined quantum updates; general learned gates require the correction definition below.

3.2 Evidence correction with unitary evolution

The subset basis of Section 3.1 also represents evidence during general unitary evolution. A pure state has the form

$$|\psi\rangle = \sum_{A\subseteq\Omega} \sqrt{m(A)}e^{\mathrm{i}\phi(A)}|A\rangle, \qquad \rho = |\psi\rangle\langle\psi|, \tag{4}$$

where the canonical MFQS has $\phi(A) = 0$. The Born masses retain their focal set meanings, while the relative phases determine how subsequent gates mix their amplitudes. A density matrix ρ covers both pure and mixed states, with the decomposition

$$m_\rho(A) = \langle A|\rho|A\rangle, \qquad \Delta(\rho) = \sum_A m_\rho(A)|A\rangle\langle A|, \qquad C_\rho = \rho - \Delta(\rho). \tag{5}$$

Here $\Delta(\rho)$ retains the represented focal masses, and C_ρ contains the off-diagonal coherence.

Definition 1 (Evidence correction with unitary evolution)**.** Let ρ_0 represent the current evidence $m_0(A) = \langle A|\rho_0|A\rangle$ in the subset basis; for the MFQS of Eq. (1), $\rho_0 = |m_0\rangle\langle m_0|$. A unitary U induces the correction $\rho_1 = U\rho_0U^\dagger$, with output evidence

$$m_1(B) = \langle B|U\rho_0U^\dagger|B\rangle = \sum_{A\subseteq\Omega} K_U(B \mid A)m_0(A) + I_U(B;\rho_0). \tag{6}$$

Writing $U_{BA} = \langle B|U|A\rangle$, the transport kernel and coherent correction are

$$K_U(B \mid A) = |U_{BA}|^2, \qquad I_U(B;\rho) = \langle B|UC_\rho U^\dagger|B\rangle. \tag{7}$$

The first term in Eq. (6) transports support between focal sets; the second accounts for interference between focal paths.This representation applies to a gate or a complete learned circuit and their derivations are given in Appendix A.

For a module G targeting ω_t, let $A \subseteq \Omega\setminus\{\omega_t\}$ and $B = A\cup\{\omega_t\}$. If G mixes only these focal pairs, put $p = m_0(A)$ and $q = m_0(B)$. Its local correction is represented by the signed flow

$$\xi_G(A) = m_1(B) - q, \qquad m_1(A) = p - \xi_G(A), \quad m_1(B) = q + \xi_G(A). \tag{8}$$

Figure 2 illustrates this representation with $\omega_t = \omega_3$. Circuit structure selects the affected pairs, while the current state and gate parameters determine their support flows. The corresponding gate formulas and their derivations are given in Appendix A.2.

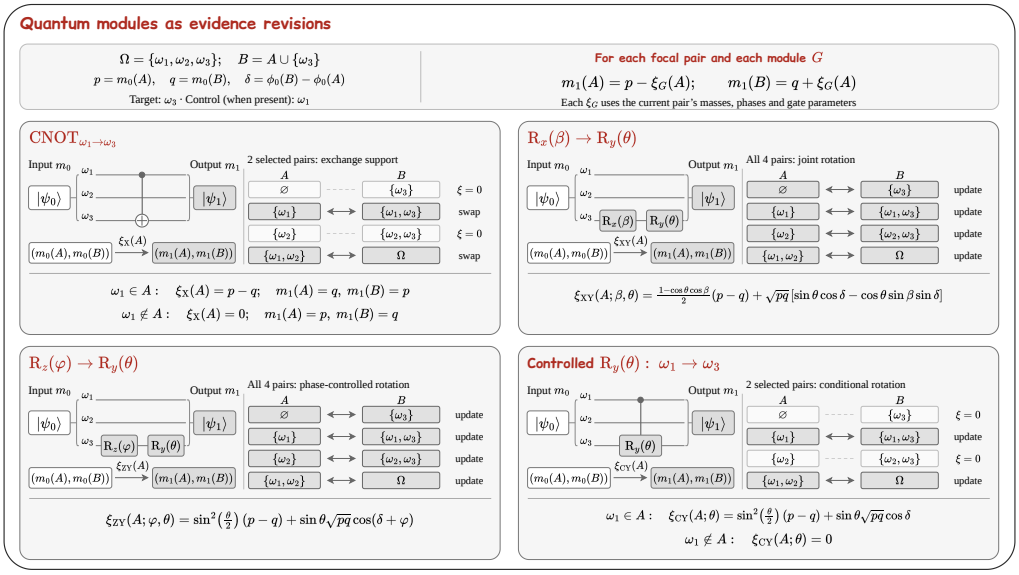

Figure 2: Examples of evidence correction with unitary evolution for pure inputs.

3.3 Properties of unitary evidence correction

The correction representation has four structural properties. Their proofs and further consequences are given in Appendix A.3.

Theorem 2 (Validity and conservation of evidence correction)**.** *For every density matrix ρ and unitary U, Eq. (6) holds. The kernel K_U is doubly stochastic, the coherent increment is real and satisfies $\sum_B I_U(B;\rho) = 0$ and the combined output is a nonnegative, normalized BPA.*

Theorem 3 (Exact composition)**.** *For successive corrections generated by $U_1, \ldots, U_L$, let $\rho_\ell = U_\ell \rho_{\ell-1} U_\ell^\dagger$. Retaining the complete state gives*

$$\rho_L = (U_L \cdots U_1)\rho_0(U_L \cdots U_1)^\dagger. \tag{9}$$

Consequently, the final evidence $m_L(B) = \langle B|\rho_L|B\rangle$ agrees exactly with the complete circuit's Born probabilities.

Proposition 4 (When the BPA alone determines an update)**.** *For every input BPA, the output BPA of a unitary is independent of all input relative phases if and only if it is a monomial unitary $U = DP$, where D is diagonal with unit-modulus entries and P is a permutation.*

Proposition 5 (Marginal-evidence preservation)**.** *Suppose a module G is block diagonal on pairs A and $B = A \cup \{\omega_t\}$, with a two-dimensional unitary on each pair. It preserves the complete non-target marginal:*

$$m_1^{\downarrow\Omega\setminus\{\omega_t\}}(A) = m_1(A) + m_1(B) = m_0(A) + m_0(B) = m_0^{\downarrow\Omega\setminus\{\omega_t\}}(A). \tag{10}$$

For $p = m_0(A)$ and $q = m_0(B)$, the flow satisfies $-q \le \xi_G(A) \le p$. The corresponding commonality change is

$$q_{m_1}(T) - q_{m_0}(T) = \begin{cases} 0, & \omega_t \notin T, \\ \displaystyle\sum_{A \supseteq T\setminus\{\omega_t\}} \xi_G(A), & \omega_t \in T, \end{cases} \tag{11}$$

where $T \subseteq \Omega$ and the sum runs over $A \subseteq \Omega \setminus \{\omega_t\}$.

For nonempty A, positive flow supports more unresolved alternatives and negative flow makes a more specific commitment; flow from $A = \varnothing$ releases conflict. The marginal invariant applies to the modules in Figure 2 and to the parity modules used by QEGN.

4 QUANTUM EVIDENTIAL GENERATIVE NETWORK

4.1 OVERVIEW

Figure 3 summarizes QEGN as an uncertainty quantification head attached to a classical classifier. For an input x with hypothesis frame $\Omega = \{\omega_1, \ldots, \omega_n\}$, the classifier supplies bounded channel tendencies $p(x)$ and compact class-conditioned activation descriptors $a(x)$. The classical network learns the input-to-class mapping, while the quantum head organizes these outputs into evidence over subsets of Ω. Singleton contours describe the tendency toward each hypothesis; composite focal sets describe alternatives that remain unresolved jointly.

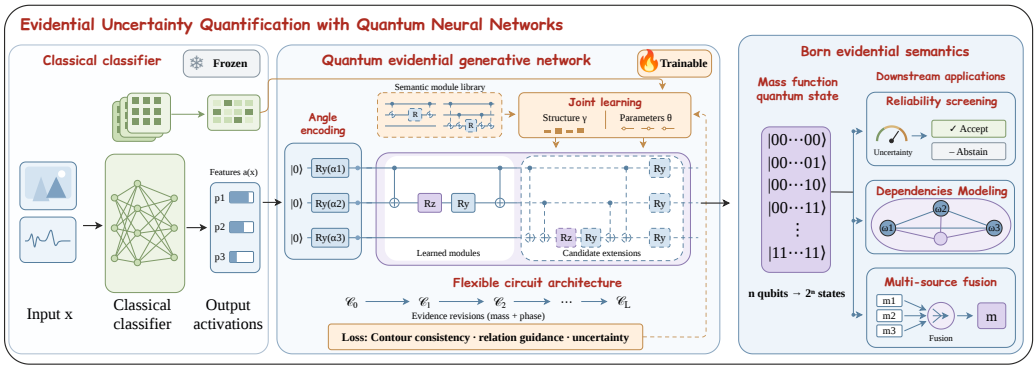

Figure 3: **Overview of QEGN.** A frozen classical classifier provides tendencies $p(x)$ for angle encoding and activation descriptors $a(x)$ for sample-dependent gate conditioning. A semantic module library supplies the circuit candidates; structure γ and parameters $\boldsymbol{\theta}$ determine the quantum head. Its successive mass-and-phase representations $\mathcal{C}_0, \ldots, \mathcal{C}_L$ have the evidence-correction interpretation of Section 3.2.

4.2 ENCODING AND LEARNING OBJECTIVE

Efficient encoding of classical tendencies. For $p_i(x) \in [0,1]$, set $\alpha_i(x) = 2 \arcsin \sqrt{p_i(x)}$ and encode

$$
\begin{aligned}
m_0(A \mid x) &= \prod_{i:\omega_i \in A} p_i(x) \prod_{i:\omega_i \notin A} [1 - p_i(x)], \\
|\psi_0(p(x))\rangle &= \bigotimes_{i=1}^{n} R_y(\alpha_i(x))|0\rangle = \bigotimes_{i=1}^{n} \left(\sqrt{1-p_i(x)}|0\rangle + \sqrt{p_i(x)}|1\rangle \right).
\end{aligned}
\tag{12}
$$

This Poss-transferable mass has independent hypothesis memberships and singleton contours $\pi_i(m_0) = p_i(x)$, which need not sum to one. Exactly n single-qubit rotations provide a linear-cost interface to 2^n subset indices.

Three complementary learning criteria. With $\bar{\pi}(m) = \pi(m)/\max_i \pi_i(m)$ and $\bar{p} = p/\max_i p_i$, QEGN minimizes

$$
\mathcal{L}_{\text{QEGN}} = \lambda_c \mathcal{L}_{\text{contour}} + \lambda_r \mathcal{L}_{\text{relation}} - \lambda_f \mathcal{H}_{\text{FB}}, \qquad \lambda_c, \lambda_r, \lambda_f > 0. \tag{13}
$$

Contour consistency: preserve relative tendencies.

$$
\mathcal{L}_{\text{contour}} = \mathbb{E}_x \, n^{-1} \|\bar{\pi}(m(x)) - \bar{p}(x)\|_2^2. \tag{14}
$$

The normalized contour softly links quantum evidence to classical relative plausibility, allowing focal masses to change while directing the QNN toward channel interrelationships rather than relearning singleton tendencies.

Relation guidance: organize unresolved alternatives. For frozen-classifier prediction $\widehat{y}(x)$, estimate $r_{ij} = \Pr(Y = j \mid \widehat{Y} = i)$ on the designated relation-calibration split and define, for $j \neq i$,

$$d_{ij} = \frac{r_{ij}}{\sum_{k\neq i} r_{ik}}, R_{ij}(m) = \sum_{A\supseteq\{\omega_i,\omega_j\}} \frac{m(A)}{|A|-1}, \widehat{d}_{ij}(m) = \frac{R_{ij}(m)}{\sum_{k\neq i} R_{ik}(m)}. \tag{15}$$

Composite focal sets distribute their mass among the anchor's other members, giving $\sum_{j\neq i} R_{ij} = \pi_i - m(\{\omega_i\})$ and measuring unresolved alternative support around ω_i. Use

$$\mathcal{L}_{\text{relation}} = -\mathbb{E}_x \sum_{j\neq\widehat{y}(x)} d_{\widehat{y}(x),j} \log \widehat{d}_{\widehat{y}(x),j}(m(x)). \tag{16}$$

Only calibration-supported rows contribute. The compact target identifies alternatives deserving joint support, leaving full focal-mass allocation to QEGN. Earlier studies use related perturbation and multimodal-rescue cues (Appendix B.3).

Fractal-based belief entropy: least commitment principle. For nonempty A, define $m^+(A) = m(A)/(1-m(\varnothing))$. The fractal-based belief (FB) transform redistributes masses over their nonempty subsets (Zhou & Deng, 2022):

$$\widetilde{m}(F) = \sum_{A\supseteq F} \frac{m^+(A)}{2^{|A|}-1}, F \neq \varnothing, H_{\text{FB}}(m) = -\sum_{F\neq\varnothing} \widetilde{m}(F)\log\widetilde{m}(F), \mathcal{H}_{\text{FB}} = \mathbb{E}_x \frac{H_{\text{FB}}(m(x))}{\log(2^n-1)}. \tag{17}$$

Minimizing negative FB entropy favors less committed evidence under contour and relation constraints. Vacuous evidence maximizes normalized entropy; the other terms constrain allocation for each sample. Empty-set mass is recorded separately. Together, these criteria learn structured uncertainty from classical outputs and compact relational guidance without complete BPA targets. The latest generator comparison trains with exactly Eq. (13); recorded task-specific extensions for earlier relation and fusion studies remain in Appendix B.3.

4.3 Circuit Generation and Parameter Optimization

A semantic module library. A structure γ orders oriented supports (S_ℓ, t_ℓ), with target $t_\ell \in S_\ell$ and context $C_\ell = S_\ell \setminus \{t_\ell\}$. Each module computes context parity with CNOTs, applies an R_z–R_y pair, and uncomputes the parity (unitary: Eq. (46); focal-pair correction: Appendix A.13). Module selection thus specifies the hypothesis contexts used for evidence correction.

Structure is fixed per trained model, while angles depend on the sample. For mean control activation $\bar{a}_{C_\ell}$, define $h_\ell = [a_{t_\ell}, \bar{a}_{C_\ell}, a_{t_\ell} - \bar{a}_{C_\ell}]$. Shared low-rank projections with module-specific outputs give bounded base-angle corrections:

$$\theta_\ell(x) = \theta_\ell^0 + s_\theta \tanh f_\ell^\theta(h_\ell(x)), \qquad \varphi_\ell(x) = \varphi_\ell^0 + s_\varphi \tanh f_\ell^\varphi(h_\ell(x)). \tag{18}$$

Parameters $\boldsymbol{\theta}$ comprise base angles and conditioner weights. The modules precede n bounded, contour-conditioned R_y adjustments; optimization differentiates the evidential objective through simulated quantum evolution.

Generate, fit, and refine a circuit. The procedure **(1)** specifies the module library and classical interface, **(2)** learns module choices and continuous parameters, **(3)** exports an ordered deterministic circuit for parameter fitting, and **(4)** optionally prunes low-response modules under contour- and relation-preservation validation criteria. The earlier response-based generator scores deletions after fitting a general circuit, then trains the selected structure from fresh initialization.

We adapt sampled score-function, hard-Gumbel, and density-matrix search to the evidential objective (Zhang et al., 2022; Chen et al., 2026a; Afane et al., 2026; Kumar et al., 2025). In the controlled four-channel comparison, each uses the same 24 two-/three-channel candidates and exports 12 modules with target coverage, yielding 740-parameter heads. Main-comparison fitting inherits exported search parameters; validation and final comparisons evaluate exported deterministic circuits. Appendix B.5 provides full algorithms and optimization settings.

For L relation modules, logical gate counts including preparation and final adjustments are

$$N_{\mathrm{CNOT}} = 2\sum_{\ell=1}^{L}(|S_\ell| - 1), \qquad N_{R_y} = 2n + L, \qquad N_{R_z} = L. \tag{19}$$

Generation adapts the correction structure while preserving the common encoding, learning objective, and evidence semantics.

5 Experiments

5.1 Learning channel relations with compact constraints

Six COVID-4/MVRSD-4 backbone configurations (He et al., 2016; Huang et al., 2017; Liu et al., 2021) test relation ordering against held-out confusion (Figure 4a, Appendix B). QEGN's mean Spearman is **0.798**, versus -0.411 with shuffled guidance and -0.664 for the bootstrap control; all six QEGN correlations are positive. Strongest-alternative recovery is **79.17%** of anchor classes, versus 16.67%/12.50% for the controls. These results support meaningful relations and class-level alternatives.

Figure 4b (Appendix B) tests prediction preservation. Contour P95 is the 95th percentile of the maximum per-sample difference between normalized input and output contours. Its six-task mean is **0.0673**, and **99.70%** of outputs retain the original top class: learned relations preserve the classifier's principal tendencies.

CIFAR-10 (Krizhevsky, 2009) uses 45 pair targets to guide a distribution over **1,024** subsets. Exact-pair mass concerns the two-class focal set; commonality includes larger sets containing that pair. Their correlations with symmetric confusion are 0.823 ± 0.040 and 0.822 ± 0.033 (Figure 4c, Appendix B), showing relation learning without full mass supervision. Accuracy is $93.237 \pm 0.055\%$, versus the backbone's 93.250%, supporting richer evidence with little predictive change.

5.2 Evidence improves multimodal decisions

Table 1 tests optical–SAR fusion at Munich and Berlin (Bahmanyar et al., 2018) and audiovisual fusion on CREMA-D (Cao et al., 2014). Full-test accuracy measures final scene/emotion decisions. Residual QEGN reaches **83.131%**, **83.888%**, and **75.196%**: mean gains of 0.192, 0.220, and 0.257 percentage points over the strongest evaluated classical rule or unimodal branch, supporting useful evidence in all three tasks.

Recovery isolates modality disagreements with exactly one correct branch. Berlin RY-CCR resolves **76.746%**: mean gains of 1.800 points over optical and 2.037 over Poss-CCR, positive in every paired seed. It also gives the best Berlin accuracy, **83.966%**. Residual QEGN gives the highest mean recovery in Munich and CREMA-D. These gains show improved conflict resolution alongside the smaller full-test gains.

Table 1: Multimodal accuracy and disagreement recovery (%), three seeds. Unimodal 1/2 denote SAR/optical or audio/video; procedure definitions are in Appendix B.4.

	Full-test accuracy ↑			Disagreement recovery ↑		
Method	**Munich ↑**	**Berlin ↑**	**CREMA-D ↑**	**Munich ↑**	**Berlin ↑**	**CREMA-D ↑**
Unimodal 1	$69.569_{\pm 0.487}$	$73.908_{\pm 0.802}$	$63.260_{\pm 1.615}$	$20.990_{\pm 0.670}$	$25.054_{\pm 1.524}$	$46.539_{\pm 1.589}$
Unimodal 2	$\underline{82.940}_{\pm 0.310}$	$83.559_{\pm 0.109}$	$65.911_{\pm 2.863}$	$\underline{79.010}_{\pm 0.670}$	$74.946_{\pm 1.524}$	$53.461_{\pm 1.589}$
P-Avg	$82.272_{\pm 0.228}$	$83.546_{\pm 0.192}$	$74.893_{\pm 2.347}$	$75.480_{\pm 0.472}$	$74.495_{\pm 0.441}$	$75.746_{\pm 1.304}$
Poss-CCR	$82.184_{\pm 0.303}$	$83.571_{\pm 0.174}$	$74.585_{\pm 2.615}$	$75.243_{\pm 0.901}$	$74.709_{\pm 0.465}$	$75.070_{\pm 0.689}$
IsoPoss-CCR	$82.126_{\pm 0.286}$	$83.668_{\pm 0.279}$	$74.940_{\pm 3.124}$	$75.004_{\pm 0.871}$	$75.244_{\pm 0.881}$	$76.268_{\pm 1.752}$
QEGN + CCR	$82.122_{\pm 0.295}$	$83.667_{\pm 0.266}$	$74.991_{\pm 3.069}$	$74.972_{\pm 0.885}$	$75.216_{\pm 0.850}$	$76.268_{\pm 1.752}$
QEGN + RY-CCR	$82.629_{\pm 0.171}$	$\mathbf{83.966}_{\pm 0.139}$	$74.890_{\pm 2.776}$	$77.192_{\pm 0.353}$	$\mathbf{76.746}_{\pm 0.219}$	$76.147_{\pm 1.317}$
QEGN + residual	$\mathbf{83.131}_{\pm 0.303}$	$\underline{83.888}_{\pm 0.091}$	$\mathbf{75.196}_{\pm 2.876}$	$\mathbf{79.680}_{\pm 0.532}$	$\underline{76.419}_{\pm 1.337}$	$\mathbf{76.941}_{\pm 0.918}$
QEGN-selected	$82.629_{\pm 0.171}$	$83.872_{\pm 0.281}$	$\underline{75.042}_{\pm 3.037}$	$77.192_{\pm 0.353}$	$76.264_{\pm 0.617}$	$\underline{76.540}_{\pm 1.605}$

Agreement with human uncertainty. Table 2 compares predictions with audiovisual vote frequencies: NLL measures cross-entropy and Brier summed squared probability differences. RY-CCR attains **1.3808** and **0.2391**, versus probability averaging's 1.4078 and 0.2463. Mean reductions are **1.91%/2.91%**; both improvements hold in every paired seed, supporting closer uncertainty allocation to human judgments, separately from acted-label accuracy.

Table 2: CREMA-D human-vote fit (lower is better), with 241/252/236 conflicting clips across three seeds.

Method	Human NLL ↓	Human Brier ↓	Method	Human NLL ↓	Human Brier ↓
Audio	$1.7750_{\pm 0.1425}$	$0.3847_{\pm 0.0335}$	Video	$1.5494_{\pm 0.1302}$	$0.3012_{\pm 0.0568}$
P-Avg	$1.4078_{\pm 0.0585}$	$0.2463_{\pm 0.0250}$	Poss-CCR	$1.4597_{\pm 0.0250}$	$0.2627_{\pm 0.0244}$
IsoPoss-CCR	$1.4537_{\pm 0.0164}$	$0.2590_{\pm 0.0206}$			
QEGN + CCR	$1.4532_{\pm 0.0157}$	$0.2587_{\pm 0.0205}$	QEGN + RY-CCR	$\mathbf{1.3808} * \pm 0.0493$	$\mathbf{0.2391} * \pm 0.0293$
QEGN + residual	$\underline{1.3916} * \pm 0.0681$	$0.2423 * \pm 0.0344$	QEGN-selected	$1.3917_{\pm 0.0681}$	$\underline{0.2421}_{\pm 0.0346}$

5.3 Uncertainty relative to EDL across three generators

Three generators share one frozen MVRSD-4 classifier and evidence objective and are compared with EDL (Sensoy et al., 2018) and DAEDL (Yoon & Kim, 2024). In Figure 5 (Appendix B), coverage is the proportion retained by uncertainty ranking and risk its error rate. AURC averages all prefix risks; lower is better. QEGN AURC is **2.6815–2.7859**, versus EDL's 3.3361 (**16.54–19.62%** lower). Accuracy among the retained 90% is **93.926–93.937%**, versus 91.490%. All nine heads improve both metrics over both tested EDL systems (Table 3).

Table 3: MVRSD-4/ResNet-34. Percentages except NLL and AURC $\times 100$. Alternatives use 335 common CE errors. QEGN uses three quantum seeds; classical systems one.

Method	Acc. ↑	AURC ↓	Acc90 ↑	Alt. NLL ↓	Alt. Top-1 ↑
EDL	85.767	3.3361	91.490	0.9174	62.99
DAEDL	$\underline{87.276}$	3.3333	92.724	0.9657	55.82
QEGN–Sampled	**90.464**	$\underline{2.7844} * \pm 0.0560$	$\mathbf{93.937} * \pm 0.156$	$\underline{0.6216} * \pm 0.0230$	$\underline{76.22} * \pm 2.32$
QEGN–Gumbel	**90.464**	$2.7859_{\pm 0.0451}$	$\underline{93.926} * \pm 0.032$	$\mathbf{0.6033} * \pm 0.0134$	$\mathbf{76.62} * \pm 0.62$
QEGN–RhoDARTS	**90.464**	$\mathbf{2.6815} * \pm 0.1793$	$\mathbf{93.937} * \pm 0.018$	$0.6535 * \pm 0.0621$	$75.12_{\pm 5.60}$

Local alternatives use the same **335** CE errors, excluding the erroneous anchor. Lower NLL means more probability on the true alternative; Top-1 counts its recovery as the most likely remaining class. QEGN reaches **0.6033–0.6535** and **75.12–76.62%**, versus EDL's 0.9174 and 62.99%, supporting more informative alternatives across generators. QEGN inherits CE's 90.464% accuracy; EDL has higher error AUROC (0.8786 versus 0.8470–0.8484). The advantages concern selective performance over these EDL systems and structured alternatives. Training and evaluation settings are given in Appendix B.5.

6 Discussion and Conclusion

This work revisits the relationship between classical and quantum neural networks through their complementary roles in uncertain information processing. QEGN uses classical networks to learn predictive tendencies and quantum networks to organize the unresolved relations among output hypotheses. Belief function theory connects these roles through singleton contours and evidence over the power set, giving the quantum state representation a concrete uncertainty semantics. The experiments illustrate how the resulting structured evidence supports relation analysis, multimodal fusion, and uncertainty assessment. Future work will explore implementations on quantum hardware and the incorporation of expert knowledge through explicitly defined evidence correction modules. QEGN provides a basis for designing quantum–classical learning systems around meaningful and complementary information roles.

AI USE STATEMENT

AI assisted with the initial drafting of experimental code, supplementary verification of proofs in the appendix, preliminary research on related work, and language polishing.

REPRODUCIBILITY STATEMENT

Appendix A provides the evidence-correction derivations, proofs, operator extensions, and worked examples, including the three-qubit construction in Appendix A.14. Appendix B specifies the data roles, objectives, training and architecture-search settings, fusion procedures, and evaluation definitions for Section 5. The accompanying artifact folder contains executable figure code and its source tables, with an index to the retained training code and checkpoints.

REFERENCES

Mohamed Afane, Quanjiang Long, Haoting Shen, Ying Mao, Junaid Farooq, Ying Wang, and Juntao Chen. Differentiable architecture search for adversarially robust quantum computer vision. *Quantum Machine Intelligence*, 8(1):2, 2026. doi: 10.1007/s42484-026-00353-0. URL https://link.springer.com/article/10.1007/s42484-026-00353-0.

Reza Bahmanyar, Daniela Espinoza-Molina, and Mihai Datcu. Multisensor earth observation image classification based on a multimodal latent dirichlet allocation model. *IEEE Geoscience and Remote Sensing Letters*, 15(3):459–463, 2018. doi: 10.1109/LGRS.2018.2794511.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. doi: 10.1038/nature23474.

Sima E. Borujeni, Saideep Nannapaneni, Nam H. Nguyen, Elizabeth C. Behrman, and James E. Steck. Quantum circuit representation of bayesian networks. *Expert Systems with Applications*, 176:114768, 2021. doi: 10.1016/j.eswa.2021.114768.

Houwei Cao, David G. Cooper, Michael K. Keutmann, Ruben C. Gur, Ani Nenkova, and Ragini Verma. CREMA-D: Crowd-Sourced Emotional Multimodal Actors Dataset. *IEEE Transactions on Affective Computing*, 5(4):377–390, 2014. doi: 10.1109/TAFFC.2014.2336244.

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(9):625–644, 2021. doi: 10.1038/s42254-021-00348-9.

M. Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J. Coles. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, 2(9):567–576, 2022. doi: 10.1038/s43588-022-00311-3.

Mengyuan Chen, Junyu Gao, and Changsheng Xu. R-EDL: Relaxing Nonessential Settings of Evidential Deep Learning. In *International Conference on Learning Representations*, pp. 35563–35588, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/98f8c89ae042c512e6c87e0e0c2a0f98-Abstract-Conference.html.

Jianpeng Chen, Zirui Sheng, Cunxi Gong, and Weitang Li. Automated Unitary Coupled Cluster Circuit Design via Differentiable Quantum Architecture Search. *arXiv preprint arXiv:2605.28049*, 2026a. doi: 10.48550/arXiv.2605.28049. URL https://arxiv.org/abs/2605.28049.

Chuangtao Chen, Qinglin Zhao, MengChu Zhou, Zhimin He, Zhili Sun, and Haozhen Situ. Quantum Generative Diffusion Model: A Fully Quantum-Mechanical Model for Generating Quantum State Ensemble. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Early Access, pp. 1–15, 2026b. doi: 10.1109/TPAMI.2026.3718311.

Xiaolong Chen, Fuyuan Xiao, Xiaohong Zhang, Zehong Cao, and Chin-teng Lin. TrustworthyQENN: A Quantum Evidential Neural Network Based on Complex-Valued Contrastive Learning for Uncertainty Pattern Classification. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026c. URL https://icml.cc/virtual/2026/poster/64299.

Koki Chinzei, Shinichiro Yamano, Quoc Hoan Tran, Yasuhiro Endo, and Hirotaka Oshima. Trade-off between gradient measurement efficiency and expressivity in deep quantum neural networks. *npj Quantum Information*, 11(1):79, 2025. doi: 10.1038/s41534-025-01036-7.

Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019. doi: 10.1038/s41567-019-0648-8.

Danruo Deng, Guangyong Chen, Yang Yu, Furui Liu, and Pheng-Ann Heng. Uncertainty Estimation by Fisher Information-based Evidential Deep Learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7596–7616, 2023. URL https://proceedings.mlr.press/v202/deng23b.html.

Thierry Denœux. Logistic regression, neural networks and Dempster–Shafer theory: A new perspective. *Knowledge-Based Systems*, 176:54–67, 2019. doi: 10.1016/j.knosys.2019.03.030.

Fan Fan, Yilei Shi, Mihai Datcu, Bertrand Le Saux, Luigi Iapichino, Francesca Bovolo, Silvia Liberata Ullo, and Xiao Xiang Zhu. Quantum circuit-based learning models: bridging quantum computing and machine learning. *Artificial Intelligence Review*, 59(10):221, 2026. doi: 10.1007/s10462-026-11686-4.

Junyu Gao, Mengyuan Chen, Liangyu Xiang, and Changsheng Xu. A Comprehensive Survey on Evidential Deep Learning and its Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):2118–2138, 2026. doi: 10.1109/TPAMI.2025.3625258.

Kaitlin Gili, Mykolas Sveistrys, and Chris Ballance. Introducing nonlinear activations into quantum generative models. *Physical Review A*, 107(1):012406, 2023. doi: 10.1103/PhysRevA.107.012406.

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(7747):209–212, 2019. doi: 10.1038/s41586-019-0980-2.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90. URL https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017. URL https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html.

Linqing Huang. Combination of Information in Labeled and Unlabeled Data via Evidence Theory. *IEEE Transactions on Artificial Intelligence*, 5(5):2179–2192, 2024. doi: 10.1109/TAI.2023.3316194.

Ling Huang, Su Ruan, Pierre Decazes, and Thierry Denœux. Deep evidential fusion with uncertainty quantification and reliability learning for multimodal medical image segmentation. *Information Fusion*, 113:102648, 2025. doi: 10.1016/j.inffus.2024.102648.

Sofiene Jerbi, Lukas J. Fiderer, Hendrik Poulsen Nautrup, Jonas M. Kübler, Hans J. Briegel, and Vedran Dunjko. Quantum machine learning beyond kernel methods. *Nature Communications*, 14(1):517, 2023. doi: 10.1038/s41467-023-36159-y.

Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Shireen Kudukkil Manchingal, Muhammad Mubashar, Kaizheng Wang, Keivan Shariatmadar, and Fabio Cuzzolin. Random-Set Neural Networks. In *International Conference on Learning Representations*, pp. 14567–14610, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/260110399ec17409228521208086c113-Abstract-Conference.html.

Swagat Kumar, Jan-Nico Zaech, Colin Michael Wilmott, and Luc Van Gool. RhoDARTS: Differentiable Quantum Architecture Search with Density Matrix Simulations. *arXiv preprint arXiv:2506.03697*, 2025. doi: 10.48550/arXiv.2506.03697. URL https://arxiv.org/abs/2506.03697v2. Version 2.

Changbin Li, Kangshuo Li, Yuzhe Ou, Lance M. Kaplan, Audun Jøsang, Jin-Hee Cho, Dong Hyun Jeong, and Feng Chen. Hyper Evidential Deep Learning to Quantify Composite Classification Uncertainty. In *International Conference on Learning Representations*, pp. 8826–8860, 2024a. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/24607f9ecc518604cb1ab03297324593-Abstract-Conference.html.

Wei Li, Deqiang Han, Jean Dezert, and Yi Yang. Basic Belief Assignment Determination Based on Radial Basis Function Network. *Chinese Journal of Information Fusion*, 1(3):175–182, 2024b. doi: 10.62762/CJIF.2024.841250.

Yawei Li, David Rügamer, Bernd Bischl, and Mina Rezaei. Calibrating LLMs with Information-Theoretic Evidential Deep Learning. In *International Conference on Learning Representations*, pp. 4298–4324, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/0cbdd1e6613c42fe975337671790f406-Abstract-Conference.html.

Jin-Guo Liu and Lei Wang. Differentiable learning of quantum circuit Born machines. *Physical Review A*, 98(6):062324, 2018. doi: 10.1103/PhysRevA.98.062324.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html.

Guang Hao Low, Theodore J. Yoder, and Isaac L. Chuang. Quantum inference on Bayesian networks. *Physical Review A*, 89(6):062315, 2014. doi: 10.1103/PhysRevA.89.062315.

Hao Luo, Qianli Zhou, Lipeng Pan, Zhen Li, and Yong Deng. Attribute fusion-based evidential classifier on quantum circuits. *Quantum Machine Intelligence*, 8(1):51, 2026a. doi: 10.1007/s42484-026-00395-4.

Hao Luo, Zhiyuan Zhai, Qianli Zhou, Jun Qi, Yong Deng, and Xin Wang. Evidential Quantum Vertical Federated Learning. *arXiv preprint arXiv:2603.21102*, 2026b. doi: 10.48550/arXiv.2603.21102. URL https://arxiv.org/abs/2603.21102.

Hailan Ma, Gary J. Mooney, Ian R. Petersen, Lloyd C. L. Hollenberg, and Daoyi Dong. Quantum autoencoders using mixed reference states. *npj Quantum Information*, 10:86, 2024. doi: 10.1038/s41534-024-00872-3.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential Deep Learning to Quantify Classification Uncertainty. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/a981f2b708044d6fb4a71a1463242520-Abstract.html.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. URL https://press.princeton.edu/books/paperback/9780691100425/a-mathematical-theory-of-evidence.

Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994. doi: 10.1016/0004-3702(94)90026-4.

Xiaoyan Su, Jian Zhong, Ziying Hong, Hong Qian, and Danilo Pelusi. A novel belief entropy and its application in cooperative situational awareness. *Expert Systems with Applications*, 286:128027, 2025. doi: 10.1016/j.eswa.2025.128027.

Yabo Wang, Xin Wang, Bo Qi, and Daoyi Dong. Supervised-learning guarantee for quantum AdaBoost. *Physical Review Applied*, 22(5):054001, 2024. doi: 10.1103/PhysRevApplied.22.054001.

Yu Wu, Qianli Zhou, Jie Geng, Xinyang Deng, and Wen Jiang. Feature Entanglement-based Quantum Multimodal Fusion Neural Network. *arXiv preprint arXiv:2601.07856*, 2026. doi: 10.48550/arXiv.2601.07856. URL https://arxiv.org/abs/2601.07856.

Fuyuan Xiao, Junhao Wen, Witold Pedrycz, and Masayoshi Aritsugi. Complex Evidence Theory for Multisource Data Fusion. *Chinese Journal of Information Fusion*, 1(2):134–159, 2024. doi: 10.62762/CJIF.2024.999646.

Fuyuan Xiao, Yu Zhou, and Witold Pedrycz. An Adaptive Quantum Circuit of Dempster's Rule of Combination for Uncertain Pattern Classification. In *Advances in Neural Information Processing Systems*, volume 38, pp. 17848–17881, 2025. doi: 10.52202/085713-0604. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/1a00115ab8033c89c01f3377bc322d9b-Abstract-Conference.html.

Li Xu, Xiao-yu Zhang, Ming Li, and Shu-qian Shen. Quantum classifiers with a trainable kernel. *Physical Review Applied*, 21(5):054056, 2024. doi: 10.1103/PhysRevApplied.21.054056.

Taeseong Yoon and Heeyoung Kim. Uncertainty Estimation by Density Aware Evidential Deep Learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 57217–57243, 2024. URL https://proceedings.mlr.press/v235/yoon24a.html.

Shi-Xin Zhang, Chang-Yu Hsieh, Shengyu Zhang, and Hong Yao. Differentiable quantum architecture search. *Quantum Science and Technology*, 7(4):045023, 2022. doi: 10.1088/2058-9565/ac87cd.

Zuowei Zhang, Liangbo Ning, Zechao Liu, Qingyu Yang, and Weiping Ding. Mining and reasoning of data uncertainty-induced imprecision in deep image classification. *Information Fusion*, 96: 202–213, 2023. doi: 10.1016/j.inffus.2023.03.014.

Ren-Xin Zhao, Jinjing Shi, and Xuelong Li. QKSAN: A Quantum Kernel Self-Attention Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10184–10195, 2024. doi: 10.1109/TPAMI.2024.3434974.

Shenkai Zhao, Xinao Zhang, Lipeng Pan, Xiaobin Xu, and Danilo Pelusi. Evidential Deep Active Learning for Semi-Supervised Classification. *arXiv preprint arXiv:2505.20691*, 2025. doi: 10.48550/arXiv.2505.20691. URL https://arxiv.org/abs/2505.20691.

Ren-Xin Zhao, Xinjie Huang, Yahong Liu, Maoyu Ye, Jinjing Shi, Shi Wang, and Yaonan Wang. QTrans: A Quantum Transformer for Sentiment Classification. *arXiv preprint arXiv:2609.12011*, 2026. doi: 10.48550/arXiv.2609.12011. URL https://arxiv.org/abs/2609.12011.

Qianli Zhou and Yong Deng. Fractal-based belief entropy. *Information Sciences*, 587:265–282, 2022. doi: 10.1016/j.ins.2021.12.032.

Qianli Zhou, Guojing Tian, and Yong Deng. BF-QC: Belief functions on quantum circuits. *Expert Systems with Applications*, 223:119885, 2023. doi: 10.1016/j.eswa.2023.119885.

Qianli Zhou, Hao Luo, Lipeng Pan, Yong Deng, and Éloi Bossé. Quantum Information Fusion and Correction under the Transferable Belief Model. *arXiv preprint arXiv:2410.08949*, 2026. doi: 10.48550/arXiv.2410.08949. URL https://arxiv.org/abs/2410.08949v6. Version 6.

A THEORY: PROOFS, EXTENSIONS, AND EXAMPLES

This appendix derives the representations stated in Section 3.2 and proves the properties in Section 3.3. In the detailed operator formulas, an index i also denotes its corresponding hypothesis ω_i, and ϕ_A abbreviates $\phi(A)$; a target written t corresponds to ω_t in Figure 2.

A.1 DERIVATION OF THE AMPLITUDE AND CORRECTION REPRESENTATIONS

For the pure-state representation in Eq. (4), the computational coordinates are $\mathcal{C} = (m, \boldsymbol{\phi})$. Define

$$\mathcal{Q}(\mathcal{C}) = \sum_{A \subseteq \Omega} \sqrt{m(A)} e^{\mathrm{i}\phi(A)} |A\rangle. \tag{20}$$

The canonical MFQS has $\phi(A) = 0$. A common phase is irrelevant, and the phase of a zero amplitude can be chosen arbitrarily. Relative phases carry the coherent context for future revisions; the BPA remains the represented uncertain information.

In these coordinates, Definition 1 is represented by $\mathsf{O}_U(\mathcal{C}_0) = \mathcal{C}_1$, where $\mathcal{C}_j = (m_j, \boldsymbol{\phi}_j)$ and $U_{BA} = \langle B|U|A\rangle$. Its complete amplitude update is

$$\begin{aligned} a_1(B) &= \sum_{A \subseteq \Omega} U_{BA} \sqrt{m_0(A)} e^{\mathrm{i}\phi_0(A)}, \qquad U_{BA} = \langle B|U|A\rangle, \\ m_1(B) &= |a_1(B)|^2, \qquad \phi_1(B) = \arg a_1(B) \quad \text{if } m_1(B) > 0. \end{aligned} \tag{21}$$

For mixed states, the same correction propagates $\rho_1 = U\rho_0 U^\dagger$, with $m_j(A) = \langle A|\rho_j|A\rangle$.

This definition extends the evidence interface beyond designated TBM combination and correction rules: U may be any learned unitary on the membership register. Its semantic description states which focal supports and coherent contexts it revises, without requiring it to coincide with a pre-existing fusion or reliability-discounting operation.

For the decomposition in Eq. (5), substituting $\rho = \Delta(\rho) + C_\rho$ into $\langle B|U\rho U^\dagger|B\rangle$ gives the transport and coherent terms defined in Eq. (7). The complete derivation and validity proof are given below.

A.2 FOCAL-PAIR CORRECTIONS ILLUSTRATED IN FIGURE 2

Consider a target hypothesis ω_t and a non-target pattern $A \subseteq \Omega \setminus \{\omega_t\}$. Its focal pair is A and $B = A \cup \{\omega_t\}$. Write

$$p = m_0(A), \quad q = m_0(B), \quad \delta = \phi_0(B) - \phi_0(A), \quad \begin{pmatrix} a_1(A) \\ a_1(B) \end{pmatrix} = e^{\mathrm{i}\phi_0(A)} V_G(A) \begin{pmatrix} \sqrt{p} \\ e^{\mathrm{i}\delta}\sqrt{q} \end{pmatrix}. \tag{22}$$

Here p, q are local pair masses, and $V_G(A)$ is the two-dimensional action of module G in that context. Squaring the output amplitudes gives the support flow in Eq. (8). Using $R_k(\gamma) = \exp(-\mathrm{i}\gamma\sigma_k/2)$, where σ_k is the Pauli operator for $k \in \{x, y, z\}$, the four modules in Figure 2 have the following rules. In the figure $\Omega = \{\omega_1, \omega_2, \omega_3\}$, the target is ω_3, and the control, when present, is ω_1:

$$\xi_{\mathrm{X}}(A) = \mathbf{1}\{\omega_1 \in A\}(p - q), \qquad G = \mathrm{CNOT}_{\omega_1 \to \omega_3}, \tag{23}$$

$$\xi_{\mathrm{ZY}}(A; \varphi, \theta) = \sin^2(\theta/2)(p - q) + \sin\theta\sqrt{pq}\cos(\delta + \varphi), \qquad G = R_y(\theta)R_z(\varphi), \tag{24}$$

$$\begin{aligned} \xi_{\mathrm{XY}}(A; \beta, \theta) = {} & \frac{1 - \cos\theta\cos\beta}{2}(p - q) \\ & + \sqrt{pq}\,[\sin\theta\cos\delta - \cos\theta\sin\beta\sin\delta], \end{aligned} \qquad G = R_y(\theta)R_x(\beta), \tag{25}$$

$$\xi_{\mathrm{CY}}(A; \theta) = \mathbf{1}\{\omega_1 \in A\}[\sin^2(\theta/2)(p - q) + \sin\theta\sqrt{pq}\cos\delta], \quad G = \text{controlled } R_y(\theta). \tag{26}$$

Products act from right to left, so $R_y(\theta)R_z(\varphi)$ denotes the figure's $R_z \to R_y$ sequence. Controlled modules select two of the four pairs; unconditioned rotations revise all four. These phase formulas describe pure inputs; for mixed inputs the corresponding real and imaginary coherence terms are obtained from ρ_{BA} through Eq. (6). The pair amplitudes in Eq. (22) also supply their output phases, which are passed to the next module. At every module boundary, m_0, m_1, p, q, δ refer to that module's current input and output.

A.3 Proofs of the correction properties

Proof of Theorem 2.

Proof. Substitute $\rho = \Delta(\rho) + C_\rho$ into $\langle B|U\rho U^\dagger|B\rangle$. The diagonal term gives $\sum_A |U_{BA}|^2 m_\rho(A)$ and the remaining term is $I_U(B;\rho)$. Unitarity gives $\sum_B |U_{BA}|^2 = \sum_A |U_{BA}|^2 = 1$, establishing both stochastic sums. Since C_ρ is Hermitian, its transformed diagonal is real. Also $\sum_B I_U(B;\rho) = \mathrm{Tr}(UC_\rho U^\dagger) = \mathrm{Tr}\, C_\rho = 0$. Finally, $U\rho U^\dagger$ is positive semidefinite with trace one, so each output mass is nonnegative and their sum is one. □

The coherent term need not be nonnegative by itself; validity applies to the complete update.

Proof of Theorem 3. For pure states, the computational coordinates reproduce the amplitude evolution through

$$\mathcal{Q}(\mathsf{O}_{U_L} \circ \cdots \circ \mathsf{O}_{U_1}(\mathcal{C}_0)) = U_L \cdots U_1 \mathcal{Q}(\mathcal{C}_0) \tag{27}$$

up to global phase. For mixed states, the same statement is Eq. (9).

Proof. Equation (21) retains both the modulus and argument of each output amplitude, so $\mathcal{Q}(\mathsf{O}_{U_\ell}(\mathcal{C})) = U_\ell \mathcal{Q}(\mathcal{C})$, up to an irrelevant common phase. Induction over ℓ proves the pure-state identity. For mixed states, $V(U\rho U^\dagger)V^\dagger = (VU)\rho(VU)^\dagger$ proves the same composition rule directly. □

Replacing an intermediate state by its positive-amplitude MFQS, or discarding its coherence, generally changes later evidence. For example, $H^2 = I$ gives $K_{H^2} = I$, whereas every entry of $K_H K_H$ is $1/2$.

Proof of Proposition 4.

Proof. For distinct focal sets A, F, consider $(|A\rangle + e^{\mathrm{i}\eta}|F\rangle)/\sqrt{2}$. The mass of output B contains $\mathrm{Re}(e^{-\mathrm{i}\eta} U_{BA} U_{BF}^*)$. Independence of every η requires $U_{BA}U_{BF}^* = 0$. Repeating for all pairs implies at most one nonzero entry in every row. Unitarity then gives exactly one entry per row and column, of modulus one, hence $U = DP$. Conversely, such a unitary only permutes masses and changes phases, so its output BPA is independent of the input phases. □

The characterization concerns the current mass update; a later mixing gate can convert phase differences into different focal masses.

Proof of Proposition 5.

Proof. Each pair's unitary preserves its squared amplitude norm, hence $m_1(A) + m_1(B) = p + q$. For a mixed state the same identity follows from trace preservation on its diagonal pair block. This sum is precisely the mass projected onto non-target pattern A. Nonnegativity of $p - \xi_G(A)$ and $q + \xi_G(A)$ gives the flow bounds. If $\omega_t \notin T$, the commonality sums complete pairs whose totals are unchanged. If $\omega_t \in T$, it sums only the target-present member of each eligible pair, whose increment is $\xi_G(A)$. □

A.4 Composition of the local operator dictionary

For a focal permutation, Eq. (42) places each complete amplitude $\sqrt{m(A)}e^{\mathrm{i}\phi_A}$ at $f(A)$, exactly as its permutation gate does. For a phase operator, Eq. (43) multiplies that amplitude by $e^{\mathrm{i}h(A)}$, exactly as the corresponding diagonal unitary does.

For a conditional revision, fix a non-target pattern D. If $g(D) = 0$, both the evidence operator and the circuit are the identity on its focal pair. If $g(D) = 1$, the fixed two-path update in Appendix A.13 is precisely $R_y(\theta)R_z(\varphi)$. Taking its squared moduli gives Eqs. (44)–(45), and retaining the output arguments gives the full output amplitudes. These disjoint pairs cover the focal basis. Hence every primitive satisfies $\mathcal{Q}(\mathsf{O}(\mathcal{C})) = U\mathcal{Q}(\mathcal{C})$. Applying this identity successively proves Eq. (27). Keeping

all focal phases permits subsequent gates to select a different target and therefore different focal pairs. Zero-amplitude phases may be chosen arbitrarily, and one global phase may be omitted. □

A.5 Pure-state coherent revision

For the phase-aware pure input in Eq. (20), the coherent increment of Theorem 2 expands as

$$I_U(B;\rho) = 2\sum_{A<C}\sqrt{m(A)m(C)}\,\mathrm{Re}\Big[U_{BA}U^*_{BC}e^{\mathrm{i}(\phi_A-\phi_C)}\Big], \tag{28}$$

where the fixed order $A < C$ counts each unordered pair once. The off-diagonal context contains interference between focal paths, not additional BPA masses.

A.6 The decomposition in a fixed focal basis

Theorem 2 proves the decomposition and its validity. For a pure state, $\rho_{AF} = \sqrt{m(A)m(F)}e^{\mathrm{i}(\phi_A-\phi_F)}$. Pairing conjugate off-diagonal terms gives Eq. (28). Once the focal basis and its dephasing map are fixed, the transport term is the output of $\Delta(\rho)$ and the coherent increment is its difference from the output of ρ; this specifies the decomposition uniquely.

A.7 Extension to quantum channels

For a completely positive trace-preserving map $\mathcal{E}$, define

$$K_{\mathcal{E}}(B \mid A) = \langle B|\mathcal{E}(|A\rangle\langle A|)|B\rangle, \qquad I_{\mathcal{E}}(B;\rho) = \langle B|\mathcal{E}(C_\rho)|B\rangle. \tag{29}$$

Then $m_{\mathcal{E}(\rho)} = K_{\mathcal{E}}m_\rho + I_{\mathcal{E}}$. Complete positivity makes $K_{\mathcal{E}}$ nonnegative, and trace preservation makes each column sum to one. Also $\sum_B I_{\mathcal{E}}(B;\rho) = \mathrm{Tr}\,\mathcal{E}(C_\rho) = 0$. These statements apply to mixed inputs and noisy circuit operations. A dephased input has $C_\rho = 0$ and therefore evolves through the stochastic focal transport term. Coherent states and classical random-set distributions are thus covered by the same focal-space formulation.

A.8 Phase information at operator boundaries

Theorem 3 retains the internal state needed by later operators. Proposition 4, proved in this appendix, characterizes the gates whose current output BPA is independent of all input phases. This does not imply that a phase gate can be discarded from a sequence: the same phase difference can be converted into different focal masses by a later mixing gate. The example derived below makes this distinction explicit.

A.9 Derivation of the membership-revision template

For an active focal pair, write $m_0 = m_-$, $m_1 = m_+$, and $\delta = \delta_D$. Use the rotation conventions

$$R_z(\varphi) = \begin{pmatrix} e^{-\mathrm{i}\varphi/2} & 0 \\ 0 & e^{\mathrm{i}\varphi/2}\end{pmatrix}, \qquad R_y(\theta) = \begin{pmatrix} c & -s \\ s & c\end{pmatrix}, \quad c = \cos(\theta/2),\ s = \sin(\theta/2). \tag{30}$$

Writing $\alpha_j = \sqrt{m_j}e^{\mathrm{i}\phi_j}$ gives

$$\alpha_0' = ce^{-\mathrm{i}\varphi/2}\alpha_0 - se^{\mathrm{i}\varphi/2}\alpha_1, \tag{31}$$

$$\alpha_1' = se^{-\mathrm{i}\varphi/2}\alpha_0 + ce^{\mathrm{i}\varphi/2}\alpha_1. \tag{32}$$

Their squared moduli yield

$$m_0' = c^2m_0 + s^2m_1 - \sin\theta\sqrt{m_0m_1}\cos(\delta+\varphi), \tag{33}$$

$$m_1' = s^2m_0 + c^2m_1 + \sin\theta\sqrt{m_0m_1}\cos(\delta+\varphi). \tag{34}$$

Subtracting m_1 gives Eq. (44) when $g(D) = 1$. Inactive pairs have zero flow, giving the same formula multiplied by $g(D)$. The arguments of the two output amplitudes give the companion phase

update; the common factor $e^{-\mathrm{i}\varphi/2}$ is retained within each active control sector. Relative to $\varphi = 0$, the additional observable effect of the phase rotation is

$$\xi_D(\varphi) - \xi_D(0) = \sin\theta\sqrt{m_0 m_1}[\cos(\delta + \varphi) - \cos\delta]. \tag{35}$$

For the canonical MFQS pair with $\delta = 0$, $\varphi = \pi/2$ removes the coherent term and $\varphi = \pi$ reverses it relative to $\varphi = 0$. Its signed magnitude also depends on θ. For example, $m_0 = m_1 = 1/2$ and $\theta = \pi/2$ give $m_1' = (1 + \cos\varphi)/2$: the same starting BPA can be transformed continuously across the complete two-outcome range through phase selection followed by mixing.

A.10 Parity modules and their conservation certificate

All CNOTs in P_C have distinct controls and the same target; they compute the control parity into that target. Their conjugation identities are

$$P_C^\dagger Z_t P_C = \left(\prod_{c\in C} Z_c\right) Z_t, \qquad P_C^\dagger Y_t P_C = \left(\prod_{c\in C} Z_c\right) Y_t. \tag{36}$$

Consequently,

$$B_{S\to t} = \exp\left[-\frac{\mathrm{i}\theta}{2}\left(\prod_{c\in C} Z_c\right) Y_t\right] \exp\left[-\frac{\mathrm{i}\varphi}{2}\left(\prod_{c\in C} Z_c\right) Z_t\right]. \tag{37}$$

On a fixed non-target membership pattern D, the product of control Z eigenvalues is ε_D. Equation (37) therefore restricts to $R_y(\varepsilon_D\theta)R_z(\varepsilon_D\varphi)$, establishing Eq. (47). The full amplitude rule, with $\bar{A} = A \triangle \{t\}$ and $\lambda_A = (-1)^{|A\cap S|}$, is

$$\alpha_A' = ce^{-\mathrm{i}\lambda_A\varphi/2}\alpha_A - \lambda_A s e^{\mathrm{i}\lambda_A\varphi/2}\alpha_{\bar{A}}. \tag{38}$$

This shows explicitly which focal amplitudes are mixed, and how the chosen support determines the parity-dependent phase and sign.

To prove Proposition 5, observe that both the conditional template and the parity module are block diagonal in every non-target membership pattern, with a two-dimensional unitary or identity on each focal pair. Summing the two output masses preserves that pair's total, which is exactly its mass after projection onto $\Omega \setminus \{t\}$. Any event involving only non-target memberships is a union of complete pairs and has invariant probability. A commonality containing t includes only the target-one member of each eligible pair, so its change is the sum of their flows. The same statement holds for mixed states by taking the trace of each diagonal two-dimensional block. More generally, any $U_S \otimes I$ preserves the marginal evidence on $\Omega \setminus S$. □

A.11 Standard evidence quantities and the gate ledger

Given a recorded increment $\Delta_\ell m(A)$, the conventional belief, plausibility, and commonality changes follow by linear summation:

$$\begin{aligned}
\Delta_\ell \,\mathrm{bel}(E) &= \sum_{\varnothing\neq A\subseteq E} \Delta_\ell m(A),\\
\Delta_\ell \,\mathrm{pl}(E) &= \sum_{A\cap E\neq\varnothing} \Delta_\ell m(A),\\
\Delta_\ell q(S) &= \sum_{A\supseteq S} \Delta_\ell m(A).
\end{aligned} \tag{39}$$

The belief definition excludes the empty set under the TBM convention. Centered moments and entropy are evaluated on each successive mass, giving $\Delta_\ell t_S = t_S(m_\ell) - t_S(m_{\ell-1})$ and $\Delta_\ell H_{\mathrm{FB}} = H_{\mathrm{FB}}(m_\ell) - H_{\mathrm{FB}}(m_{\ell-1})$. Every such difference telescopes along the executed circuit.

For each semantic module, the ledger may store

$$\mathcal{E}_\ell = (S_\ell, t_\ell, \theta_\ell, \varphi_\ell, K_\ell m_{\ell-1}, I_\ell, \Delta_\ell m, \Delta_\ell \bar{\pi}, \Delta_\ell q, \Delta_\ell t, \Delta_\ell H_{\mathrm{FB}}). \tag{40}$$

At the boundary of a complete parity module, focal labels are restored, making these records directly interpretable in the original hypothesis space. A phase-only gate can leave the current mass unchanged while modifying the context used by the next record. A CNOT has a permutation transport kernel and acts as conditional membership relabeling. These are distinct, computable operations within the same revision language.

A.12 REUSABLE LOCAL EVIDENCE-REVISION TEMPLATES

Previous TBM circuits construct quantum implementations of specified evidence operations (Zhou et al., 2026). We develop the complementary interface: a gate or gate block is translated into a fixed evidence-revision operator. Its structure specifies the affected focal sets and membership conditions; its parameters and input evidence determine the redistribution. The operator is defined once and reused across inputs, circuit positions, and frame sizes.

For pure-state circuits, the computational interface retains the mass and focal phases,

$$\mathcal{C} = (m, \boldsymbol{\phi}), \qquad \mathcal{Q}(\mathcal{C}) = \sum_A \sqrt{m(A)} e^{\mathrm{i}\phi_A} |A\rangle. \tag{41}$$

The mass has its conventional evidential meaning; the phases carry the coherent context for subsequent revisions. The canonical preparation fixes their initial values to zero, so the external input remains the initial evidence and the parameterized circuit. Three operator families provide the interface.

Focal permutation. For a bijection $f : 2^\Omega \to 2^\Omega$, define P_f by

$$m'(f(A)) = m(A), \qquad \phi'_{f(A)} = \phi_A. \tag{42}$$

It moves support without splitting it. An X_t gate uses $f(A) = A \triangle \{t\}$; CNOT applies this toggle only when the control belongs to A; Toffoli uses a conjunction of control memberships. SWAP exchanges two membership positions. These operations move mass among fixed hypothesis subsets rather than changing the hypotheses themselves.

Focal phase modulation. For a specified phase function $h(A)$, define Φ_h by

$$m'(A) = m(A), \qquad \phi'_A = \phi_A + h(A). \tag{43}$$

For $R_{zt}(\varphi)$, $h(A) = \varphi(2\mathbf{1}\{t \in A\} - 1)/2$; CZ assigns phase π exactly to sets containing both selected hypotheses. These operators preserve current support while setting the coherent conditions under which later mixing redistributes it.

Conditional membership revision. Fix a target t and a Boolean predicate $g(D)$ on the non-target membership pattern $D \subseteq \Omega \setminus \{t\}$. Define $\mathsf{R}_{t|g}(\theta, \varphi)$ as $R_{zt}(\varphi)$ followed by $R_{y_t}(\theta)$ in the sectors with $g(D) = 1$, and the identity otherwise. It acts on the nested focal pair $D, D \cup \{t\}$. Write $m_- = m(D)$, $m_+ = m(D \cup \{t\})$, and $\delta_D = \phi_{D\cup\{t\}} - \phi_D$. Every invocation uses the same rule:

$$\xi_D = g(D) \left[\sin^2 \frac{\theta}{2} (m_- - m_+) + \sin\theta \sqrt{m_- m_+} \cos(\delta_D + \varphi) \right], \tag{44}$$

$$m'(D) = m_- - \xi_D, \qquad m'(D \cup \{t\}) = m_+ + \xi_D. \tag{45}$$

Positive flow moves support toward the larger set; negative flow moves it toward the smaller set. For nonempty nested sets these are local generalization and specialization directions: retaining more alternatives or making a more specific commitment. The predicate selects *where* to revise; the input and angles determine *how much and in which direction*. The fixed phase update accompanying Eq. (45) is given in Appendix A.13.

Ordinary $R_{y_t}(\theta)$ is $\mathsf{R}_{t|1}(\theta, 0)$; controlled $R_{y_{c\to t}}(\theta)$ uses $g(D) = \mathbf{1}\{c \in D\}$. A controlled phase–rotation pair invokes the same template with nonzero φ. Other single-qubit gates have a common three-coefficient mass-flow form (Appendix A.13). Thus a new circuit is interpreted by selecting and composing rules, not by deriving a new explanation for each numerical output.

Compiling the learned relation modules. The QEGN modules used below combine these same primitives through a parity map. For $S = C \cup \{t\}$, let $P_C = \prod_{c\in C} \mathrm{CNOT}_{c\to t}$ and define

$$B_{S\to t}(\theta, \varphi) = P_C^\dagger R_{y_t}(\theta) R_{zt}(\varphi) P_C. \tag{46}$$

The two focal permutations surround one phase-modulated membership revision. In context D, put $\varepsilon_D = (-1)^{|D\cap C|}$. The block applies the same revision rule with angles $\varepsilon_D\theta, \varepsilon_D\varphi$:

$$\xi_D^{S\to t} = \sin^2 \frac{\theta}{2} (m_- - m_+) + \varepsilon_D \sin\theta \sqrt{m_- m_+} \cos(\delta_D + \varepsilon_D \varphi). \tag{47}$$

Conjunction-controlled gates select which contexts are active; parity modules select the signed revision within each context. Both obey Proposition 5. Circuit search therefore chooses the contexts and targets of evidence revision, while parameter training fits the associated support flows. The resulting interpretation is an operator-level description of the learned function, with numerical traces available as evaluations of these fixed rules.

A.13 A REUSABLE DICTIONARY OF FOCAL OPERATORS

The dictionary below instantiates the focal permutation P_f, focal phase modulation Φ_h, and conditional membership revision $\mathsf{R}_{t|g}(\theta, \varphi)$ defined in the preceding operator section. Each entry specifies an operation before an input is supplied. Its mass and phase outputs can therefore be passed directly to the next entry, as required by Theorem 3. All subsets index the same fixed frame; moving support between subsets does not rename its hypotheses.

A.13.1 FIXED COEFFICIENTS FOR ANY SINGLE-QUBIT GATE

For a target t and context $D \subseteq \Omega \setminus \{t\}$, set $A_0 = D$, $A_1 = D \cup \{t\}$, $m_- = m(A_0)$, $m_+ = m(A_1)$, and $\alpha_b = \sqrt{m(A_b)}e^{\mathrm{i}\phi_b}$. Define

$$r_D = \alpha_1\alpha_0^* = \sqrt{m_-m_+}e^{\mathrm{i}\delta_D}, \qquad \delta_D = \phi_1 - \phi_0. \tag{48}$$

For a fixed unitary $V = (v_{ab})$, store three real coefficients:

$$\begin{aligned} \lambda &= |v_{10}|^2, \qquad u = 2\,\mathrm{Re}(v_{10}^*v_{11}), \qquad v = -2\,\mathrm{Im}(v_{10}^*v_{11}), \\ \xi_D &= \lambda(m_- - m_+) + u\,\mathrm{Re}\,r_D + v\,\mathrm{Im}\,r_D, \\ (m'_-, m'_+) &= (m_- - \xi_D, m_+ + \xi_D), \\ \alpha'_b &= v_{b0}\alpha_0 + v_{b1}\alpha_1, \qquad \phi'_b = \arg\alpha'_b \quad \text{when } m'_b > 0. \end{aligned} \tag{49}$$

Here $m'_0 = m'_-$ and $m'_1 = m'_+$. Phases of zero amplitudes may be chosen arbitrarily. Expanding $|\alpha'_1|^2$ and using $|v_{11}|^2 = 1 - |v_{10}|^2$ proves the mass rule; the amplitude rule supplies its exact phase update. Thus ξ_D is a signed transfer toward the focal set containing t, with its direction determined by the input and parameters.

Table 4: Fixed single-qubit dictionary. Write $c = \cos(\theta/2)$ and $s = \sin(\theta/2)$. The last column gives the full output amplitudes, hence both focal phases by their arguments.

Gate	(λ, u, v)	(α'_0, α'_1)
$R_y(\theta)$	$(s^2, \sin\theta, 0)$	$(c\alpha_0 - s\alpha_1,\ s\alpha_0 + c\alpha_1)$
$R_x(\theta)$	$(s^2, 0, -\sin\theta)$	$(c\alpha_0 - \mathrm{i}s\alpha_1,\ -\mathrm{i}s\alpha_0 + c\alpha_1)$
H	$(1/2, -1, 0)$	$((\alpha_0 + \alpha_1)/\sqrt{2},\ (\alpha_0 - \alpha_1)/\sqrt{2})$
X	$(1, 0, 0)$	(α_1, α_0)
Y	$(1, 0, 0)$	$(-\mathrm{i}\alpha_1, \mathrm{i}\alpha_0)$
I	$(0, 0, 0)$	(α_0, α_1)
$Z, S, T, P(\varphi)$	$(0, 0, 0)$	$(\alpha_0, e^{\mathrm{i}\kappa}\alpha_1)$, with $\kappa = \pi, \pi/2, \pi/4, \varphi$, respectively
$R_z(\varphi)$	$(0, 0, 0)$	$(e^{-\mathrm{i}\varphi/2}\alpha_0, e^{\mathrm{i}\varphi/2}\alpha_1)$

A.13.2 FIXED PERMUTATION AND PHASE MAPS

A monomial gate has one nonzero entry per row and column. Its dictionary entry is the pair (f, h):

$$\begin{aligned} U|A\rangle &= e^{\mathrm{i}h(A)}|f(A)\rangle, \qquad \mathcal{C}' = (\mathsf{P}_f \circ \Phi_h)(\mathcal{C}), \\ m'(f(A)) &= m(A), \qquad \phi'_{f(A)} = \phi_A + h(A). \end{aligned} \tag{50}$$

Thus the phase in Table 5 is indexed by the *input* focal set, before permutation. Let $b_i = \mathbf{1}\{i \in A\}$, let $\oplus$ denote exclusive OR, and let σ_{ij} exchange membership positions i, j.

Here $\mathrm{RZZ}_{i,j}(\varphi) = \exp(-\mathrm{i}\varphi Z_iZ_j/2)$. The predicates in these maps refer to focal membership, not Bayesian conditioning on a true class or fusion of independent sources. A phase increment supplies coherent context for subsequent mixing; it is not itself a reliability or ignorance parameter.

Table 5: Gate-to-operator maps, valid for every coherent input. A conditional toggle is the identity when its stated condition fails. Pure permutations move each focal mass intact; pure phase maps preserve all current masses.

Gate	Focal permutation $f(A)$	Input phase $h(A)$
X_t	$A\triangle\{t\}$	0
Y_t	$A\triangle\{t\}$	$\pi/2 - \pi b_t$
Z_t, S_t, T_t	A	$(\pi, \pi/2, \pi/4)b_t$
$P_t(\varphi)$	A	φb_t
$R_{zt}(\varphi)$	A	$\varphi(b_t - 1/2)$
$\mathrm{CNOT}_{j\to t}$	$A\triangle\{t\}$ if $b_j = 1$	0
$\mathrm{Toffoli}_{j,k\to t}$	$A\triangle\{t\}$ if $b_j b_k = 1$	0
$\mathrm{CZ}_{i,j}$	A	$\pi b_i b_j$
$\mathrm{SWAP}_{i,j}$	$\sigma_{ij}(A)$	0
$\mathrm{RZZ}_{i,j}(\varphi)$	A	$\varphi[(b_i \oplus b_j) - 1/2]$

A.13.3 The complete conditional $R_z \to R_y$ template

The entry $\mathsf{R}_{t|g}(\theta, \varphi)$ applies $R_{y_t}(\theta)R_{z_t}(\varphi)$ on precisely the contexts where $g(D) = 1$ and the identity elsewhere. In an active context put

$$\begin{aligned} w_- &= c\sqrt{m_-} - se^{\mathrm{i}(\delta_D+\varphi)}\sqrt{m_+}, & w_+ &= s\sqrt{m_-} + ce^{\mathrm{i}(\delta_D+\varphi)}\sqrt{m_+}, \\ \alpha_0' &= e^{\mathrm{i}(\phi_0-\varphi/2)}w_-, & \alpha_1' &= e^{\mathrm{i}(\phi_0-\varphi/2)}w_+, \\ \phi_0' &= \phi_0 - \varphi/2 + \arg w_-, & \phi_1' &= \phi_0 - \varphi/2 + \arg w_+. \end{aligned} \tag{51}$$

The phase lines apply to nonzero outputs, modulo 2π. Their squared moduli give Eq. (44). Inactive contexts retain both masses and phases. The shared term $-\varphi/2$ in an active pair is retained because it need not be a global phase of the entire state.

More generally, let $\Pi_g = \sum_{D:g(D)=1}|D\rangle\langle D|$ act on the non-target membership coordinates. Controlled V uses the same coefficient and amplitude entry as V, only on active pairs:

$$U_{g,t}(V) = (I - \Pi_g) \otimes I_t + \Pi_g \otimes V_t. \tag{52}$$

The target tensor position is implicit. This covers conjunctions such as $g(D) = \mathbf{1}\{C \subseteq D\}$ and other fixed membership predicates. Although $P(\varphi) = e^{\mathrm{i}\varphi/2}R_z(\varphi)$ unconditionally, their controlled versions differ in the relative phase between control sectors. Likewise, CNOT and controlled-$R_y(\pi)$ both swap active-pair masses, but the latter maps (α_0, α_1) to $(-\alpha_1, \alpha_0)$; their coherent actions are different.

Pair conservation implies the non-target marginal-evidence invariant in Proposition 5. The parity module in Eq. (46) is a composition of dictionary entries; Eq. (47) supplies its context-dependent signed flow. Conjunction selects whether a pair is revised, whereas that parity construction selects rotation signs.

A.13.4 Composition, mixed states, and evidence queries

The dictionary propagates all focal phases, up to one global phase, so later targets can pair different subsets. Each entry reproduces the full amplitudes, establishing Theorem 3 by induction.

For mixed inputs, use $r_D = \rho_{A_1,A_0}$ in the same mass rule and propagate the full $\rho' = U\rho U^\dagger$, retaining cross-context coherences for later targets. The pure-amplitude phase lines are replaced by this update. A monomial entry gives $\rho'_{f(A),f(B)} = e^{\mathrm{i}[h(A)-h(B)]}\rho_{A,B}$.

The mass output determines the belief, plausibility, and commonality changes through Eq. (39).

A.14 A THREE-QUBIT EXAMPLE OF REUSABLE EVIDENCE OPERATORS

This example applies the same conditional membership-revision template twice. The circuit specifies the active focal pairs before execution; the shared flow rule in Eq. (44) then determines the amount and direction of each revision from its input. All parameters below are chosen for an exact, hand-computable illustration and are not learned parameters or experimental results.

Encoding and fixed calls. Let $\Omega = \{a, b, c\}$ and use the bit order $|abc\rangle$, so that, for example, $|101\rangle$ represents the focal set $\{a, c\}$. Starting from $|000\rangle$, apply $R_{y_a}(\pi/3)$, $R_{y_b}(\pi/3)$, and $R_{y_c}(\pi/3)$. The resulting coherent state $\mathcal{C}_0 = (m_0, \mathbf{0})$ is

$$|\psi_0\rangle = \left(\frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle\right)^{\otimes 3}, \qquad m_0(A) = \left(\frac{1}{4}\right)^{|A|}\left(\frac{3}{4}\right)^{3-|A|}. \tag{53}$$

Each initial singleton contour equals $1/4$. We retain the empty-set mass under the TBM convention, without closed-world renormalization. Consider the following gates, ordered from left to right in time:

$$\mathrm{CRY}_{a\to c}(\pi/3) \;\longrightarrow\; \mathrm{CRZ}_{b\to c}(\pi) \;\longrightarrow\; \mathrm{CRY}_{b\to c}(\pi/3). \tag{54}$$

Define $g_a(D) = \mathbf{1}_{a\in D}$ and $g_b(D) = \mathbf{1}_{b\in D}$ for $D \subseteq \{a, b\}$. The circuit is exactly the two fixed calls

$$\mathcal{C}_1 = \mathsf{R}_{c|g_a}(\pi/3, 0)(\mathcal{C}_0), \qquad \mathcal{C}_2 = \mathsf{R}_{c|g_b}(\pi/3, \pi)(\mathcal{C}_1). \tag{55}$$

Thus both calls revise membership of c, first in contexts containing a and then in contexts containing b. The second call includes its phase modulation before mixing. These are focal-set membership conditions, not conditioning on an observed true class.

Evaluating the two calls. For an active pair $(D, D \cup \{c\})$, write $m_- = m(D)$ and $m_+ = m(D \cup \{c\})$. Initially $\delta_D = 0$; the first call therefore uses

$$\xi_D^{(1)} = \frac{1}{4}(m_- - m_+) + \frac{\sqrt{3}}{2}\sqrt{m_- m_+}. \tag{56}$$

Only $D = \{a\}$ and $D = \{a, b\}$ are active. For example, $\xi_{\{a\}}^{(1)} = (9-3)/(4\cdot 64) + \sqrt{3}\sqrt{9\cdot 3}/(2\cdot 64) = 6/64$. The pair update in Eq. (45) moves this mass from $\{a\}$ to $\{a, c\}$. Both active pairs expand their candidate sets, as listed in Table 6.

All amplitudes after this first call remain positive real numbers, so the second call receives $\delta_D = 0$ on its active pairs. This is the computed output phase, not a phase reset or an intermediate measurement. Substituting $\varphi = \pi$ in the same rule gives

$$\xi_D^{(2)} = \frac{1}{4}(m_- - m_+) - \frac{\sqrt{3}}{2}\sqrt{m_- m_+}, \qquad (m_-, m_+) \text{ taken from } m_1. \tag{57}$$

Now $D = \{b\}$ and $D = \{a, b\}$ are active. Their negative flows contract support toward the smaller focal sets. In particular, the entire mass on $\{b, c\}$ moves to $\{b\}$. The shared pair $(\{a, b\}, \Omega)$ receives both revisions in sequence; the second call uses its updated input $(1, 3)/64$.

Table 6: The active focal pairs in the two calls. Every mass and flow entry is expressed in units of $1/64$; all unlisted pairs are unchanged by that call.

Call	Focal pair	Input (m_-, m_+)	Flow ξ_D	Output (m'_-, m'_+)
1	$\{a\} \leftrightarrow \{a, c\}$	$(9, 3)$	$+6$	$(3, 9)$
1	$\{a, b\} \leftrightarrow \Omega$	$(3, 1)$	$+2$	$(1, 3)$
2	$\{b\} \leftrightarrow \{b, c\}$	$(9, 3)$	-3	$(12, 0)$
2	$\{a, b\} \leftrightarrow \Omega$	$(1, 3)$	-2	$(3, 1)$

Complete output and evidence queries. Table 7 gives every mass and the associated contours and commonalities. Since both calls target c, the full marginal mass function on $\{a, b\}$ is invariant. Consequently, $\mathrm{pl}(a)$, $\mathrm{pl}(b)$, and $q(\{a, b\})$ remain unchanged, whereas quantities involving c can change. For example, $q(\{a, c\}) = m(\{a, c\}) + m(\Omega)$ increases overall, while $q(\{b, c\})$ decreases. These commonalities measure support on focal sets that retain both candidates; they are not probabilities that two mutually exclusive classes are simultaneously true.

Table 7: Exact initial, intermediate, and final masses (left, scaled by 64), and evidence queries (right, unscaled). Both calls preserve total mass, including $m(\varnothing) = 27/64$.

Basis	Focal set	$64m_0$	$64m_1$	$64m_2$
000	$\varnothing$	27	27	27
001	$\{c\}$	9	9	9
010	$\{b\}$	9	9	12
011	$\{b, c\}$	3	3	0
100	$\{a\}$	9	3	3
101	$\{a, c\}$	3	9	9
110	$\{a, b\}$	3	1	3
111	Ω	1	3	1

Quantity	Initial	Call 1	Call 2
$\mathrm{pl}(a)$	1/4	1/4	1/4
$\mathrm{pl}(b)$	1/4	1/4	1/4
$\mathrm{pl}(c)$	1/4	3/8	19/64
$q(\{a, c\})$	1/16	3/16	5/32
$q(\{b, c\})$	1/16	3/32	1/64
$q(\{a, b\})$	1/16	1/16	1/16
$q(\Omega)$	1/64	3/64	1/64

Retaining the state for further composition. The final nonzero amplitudes have phases $\phi_{\{b\}} = \phi_{\{a,b\}} = -\pi/2$ and $\phi_\Omega = \pi/2$; all other nonzero amplitudes have phase zero. The zero-mass focal set $\{b, c\}$ has no consequential phase. Keeping these phases in $\mathcal{C}_2 = (m_2, \boldsymbol{\phi}_2)$ permits exact subsequent calls, including calls with a different target that pair the focal sets differently. The example demonstrates composable operator semantics: conditional expansion followed by conditional contraction, both calculated by one fixed rule. These exact values illustrate the operator language independently of the trained-head experiments.

A.15 Preparation and simultaneous evidence queries

The Poss-transferable initialization and its rotation angles are defined in Eq. (12). Equation (12) follows by expanding the tensor product. Squaring its amplitudes gives the product mass, and summing over the other memberships gives p_i for each singleton. The preparation uses exactly n rotations, with no multi-qubit input gate.

Let $\mathcal{F}$ be a fixed collection of M queried subsets, including singleton queries if desired. Draw N independent focal samples $A^{(1)}, \ldots, A^{(N)}$ from a computational-basis measurement and set

$$\widehat{q}(S) = \frac{1}{N} \sum_{r=1}^{N} \mathbf{1}\{S \subseteq A^{(r)}\}. \tag{58}$$

For each S, this is an unbiased mean of Bernoulli variables. Hoeffding's inequality and a union bound give

$$\Pr\left(\max_{S \in \mathcal{F}} |\widehat{q}(S) - q(S)| > \epsilon \right) \leq 2Me^{-2N\epsilon^2}. \tag{59}$$

Solving for the desired failure probability gives $N \geq \log(2M/\delta)/(2\epsilon^2)$ shots for simultaneous additive error ϵ with failure probability at most δ. For all queries through order K, $M = \sum_{k=1}^{K} \binom{n}{k}$, a polynomial count for fixed K. The same samples are reused for all subset-inclusion tests. The classical aggregation work depends on the requested query list. Complete mass-vector and FB evaluations in the reported simulations use the full statevector; the shared-shot result describes direct contour and commonality queries.

For $|S| \geq 2$, centered membership relations are $t_S(m) = \mathbb{E}_m[\prod_{i \in S}(z_i - \pi_i)]$. They can be recovered algebraically from commonalities:

$$t_S = \sum_{T \subseteq S} (-1)^{|S|-|T|} q(T) \prod_{i \in S \setminus T} \pi_i, \qquad q(\varnothing) = 1. \tag{60}$$

This connects the observable relation statistics to the same focal-membership semantics. For a selected list of oriented modules, logical cost is given by Eq. (19); every repeated module is counted. Using one orientation of every support through order K gives $2\sum_{k=2}^{K}(k-1)\binom{n}{k}$ CNOTs.

B Experimental settings for Section 5

This appendix specifies the protocols used by the experiments in Section 5.

B.1 FIGURES

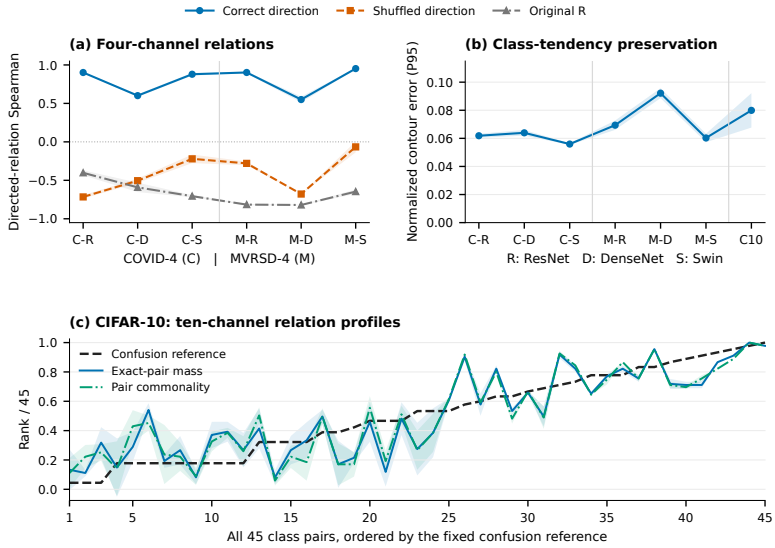

Figure 4: **Relations and prediction preservation.** (a) Direction Spearman; (b) contour P95 across task configurations; (c) normalized ranks of all 45 CIFAR-10 pairs ordered by confusion. C/M: COVID/MVRSD; R/D/S: ResNet/DenseNet/Swin. Means and SD use three quantum seeds.

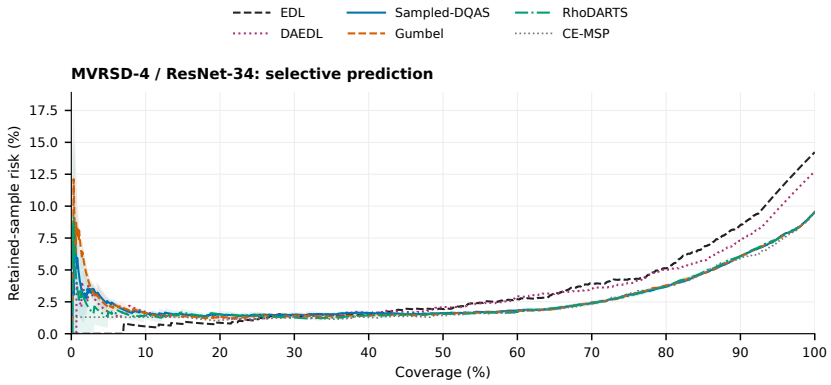

Figure 5: **MVRSD-4 selective prediction.** Full curves retain each system's decisions. Bands show three-seed SD, bounded below by zero; classical systems are single-seed.

B.2 IMPLEMENTATION AND DATA PROTOCOLS

The numerical results use the completed QEGN instantiations and experiment-family settings recorded below. Circuit modules and classical interfaces are shared across these instantiations; the recorded training criteria are specified in Appendix B.3.

B.2.1 CLASSICAL INTERFACES

For COVID-4, MVRSD-4, and CIFAR-10, let z_i be the frozen classifier logits and T its calibrated temperature. The retained interface computes

$$e_i = \mathrm{softplus}(z_i/T), \qquad u = \frac{n}{\sum_i e_i + n}, \qquad p_i = (1-u)\frac{e_i}{\max_j e_j}. \tag{61}$$

This raw contour is encoded by the Poss preparation. Tendencies are clipped to $[10^{-4}, 1-10^{-4}]$ before their preparation angles are calculated. Both input and output are independently max-normalized when calculating the consistency loss. For the multimodal tasks, the interface instead encodes $p_i = \mathrm{softmax}(z/T)_i / \max_j \mathrm{softmax}(z/T)_j$, using validation-calibrated temperatures. The two constructions use the same qubit-to-hypothesis correspondence and the same normalized-consistency criterion.

Each channel has eight activation descriptors. In the four-class experiments, its class activation map is obtained from the final spatial features and classifier weights. Subtracting its spatial mean isolates the spatial contribution, and the eight retained non-DC DCT coordinates are $(0,1), (1,0), (2,0), (1,1), (0,2), (0,3), (1,2), (2,1)$. They are standardized on the quantum-fitting pool. For CIFAR-10, class weights multiply the 512-dimensional pooled feature elementwise, followed by a fixed orthonormal projection to eight dimensions and classwise standardization. For multimodal tasks, weighted pooled features are projected by an eight-dimensional PCA fitted on the relation role, standardized, and passed through $\tanh$.

For a module (S, t), let $\bar{a}_C$ be the mean activation of its controls. Its conditioner input is $h_{S,t} = [a_t, \bar{a}_C, a_t - \bar{a}_C]$. Separate rank-four projections for the amplitude and phase angles are shared by module order. Each module then applies its own linear readout and bias, followed by the bounded $\tanh$ modulation in Section 4. The scales are $s_\theta = 0.25$ and $s_\varphi = 0.35$. The final adjustment angles are $0.20 \tanh(Wp + b)$. The head implementation uses PennyLane with a PyTorch interface and backpropagation through the statevector; equivalent batched mass operations implement the evidence transforms and fusion kernels.

B.2.2 Four-class data roles and training

Each dataset is evaluated with ResNet-34, DenseNet-121, and Swin-T. The quantum fitting pools contain 14,390 COVID-4 and 13,107 MVRSD-4 samples from the retained classifier-training cache. The original test pool supplies a fixed, class-stratified calibration/selection/held-out partition, using the same indices for all backbones:

Dataset	Confusion calibration	Structure selection	Final held-out
COVID-4	1,270	847	2,118
MVRSD-4	1,053	703	1,757

The fixed split seed is 20260824, with seed plus one for the second stratified split. New correct and shuffled heads each use three seeds, giving 36 trained heads. The retained R control adds 18 evaluated heads.

The generic library contains all 12 oriented pairs and 12 oriented triples. It is fitted for eight epochs, 16 modules are selected by their functional response with target coverage, and the selected head is initialized afresh and fitted for 24 epochs. Refinement tests one additional module deletion, with four recovery epochs when used. Optimization uses AdamW, learning rate 0.005, weight decay 10^{-4}, batch size 2,048, a cosine learning-rate schedule, gradient norm cap 1, and class-balanced resampling. Loss weights are $\lambda_s = 0.25$, $\lambda_d = 0.75$, $\lambda_o = 0.10$, and $\lambda_c = 30$. The positive FB weight follows a cosine increase from 0.02 to 0.08.

B.2.3 Ten-class data roles and structure search

The official CIFAR-10 training pool is split into 45,000 classifier-training and 5,000 validation images; its 10,000 test images are used for evaluation. The compact ResNet-9 classifier is trained for 40 epochs, with epoch 37 selected by validation accuracy. The temperature is 2.164128. The head search evaluates 24 candidates by successive halving: stages retain 8, 3, and 1 candidates while adding 1, 2, and 3 epochs and using 80, 150, and 300 training examples per class. Selection uses 100 validation examples per class.

The selected circuit contains 45 pair modules and 842 formal trainable parameters, including the activation conditioner. Final parameters are trained from fresh initialization for 20 epochs, using 500 training examples per class; the selected epochs across seeds are 13, 8, and 17. The learning rate is 0.006, weight decay 10^{-4}, pair-relation weight 10, order weight 0.03, contour weight 10, and FB

weight 0.01–0.06. The augmented contour criterion uses tolerance 0.003, penalty coefficient 40, and dual learning rate 8.

B.2.4 MULTIMODAL DATA AND CLASSIFIERS

MSLCC contains four land-cover classes: agriculture, forest, built-up, and water. Munich and Berlin are treated as separate same-domain fusion tasks. A 10×10 geographic grid is assigned to classifier/relation/fusion/validation/test roles with fractions 45/15/15/10/15%. Samples are centered inside the blocks with a 16-pixel margin for 33×33 patches. Actual per-city role sizes are 72,000/16,000/16,000/16,000/32,000, with balanced classes. Both branches use ImageNet-pretrained ResNet-18, adapted to four optical channels or one SAR channel using a 3×3 first convolution of stride one and no max pooling. Validation selects four-rotation SAR test-time augmentation for Munich and eight dihedral views for Berlin. Three training seeds share the geographic partition.

CREMA-D uses anger, disgust, fear, happiness, neutral, and sadness. Actor-disjoint roles contain 55 classifier actors, 12 relation actors, and eight actors each for fusion, validation, and test. The three partitions have the following clip counts:

Seed	Classifier	Relation	Fusion	Validation	Test
0	4,491	983	656	656	656
1	4,498	984	656	656	648
2	4,503	977	656	650	656

The audio branch is an ImageNet-pretrained ResNet-18 operating on one-channel log-mel features: 16 kHz sampling, four-second windows, 128 mel bins, FFT size 400, and hop size 160. The video branch is a Kinetics-400-pretrained R3D-18 receiving 16 frames at 112×112 pixels. The relation role is split internally 80/20 for fitting and selection. At most 4,096 samples are used for fitting each quantum candidate. Candidate module budgets are 8/12/16 for MSLCC and 16/24/32 for CREMA-D, with proxy/final training of 10/30 epochs and Adam learning rate 0.003.

B.2.5 STRUCTURE LEARNING PROCEDURE

The following procedure records the earlier response-based generator; joint-search variants are specified in Appendix B.5.

1. Fit the classical task network, freeze its feature and prediction interfaces, and extract bounded contours and class activations.
2. Construct relation observations from the designated fitting/calibration role, and initialize a generic circuit from the oriented module library.
3. Train its continuous parameters and measure the effect of replacing candidate modules by the identity on a held-out selection objective.
4. Rank candidate modules or candidate ordered circuits, enforcing target coverage. Where used, apply successive halving to concentrate training on promising candidates.
5. Fix the selected structure, discard proxy weights and optimizer state, and train a new head from fresh initialization.
6. Test low-response modules for removal. Retain a shorter circuit when the validation preservation criterion is met; apply a short recovery fit if configured.
7. Freeze the head. Query its evidence or fit the downstream fusion operator using the designated fusion role, select on validation data, and evaluate the final retained system.

For multimodal structure scoring, the module deletion response combines 0.35 times the percentile rank of evidential-loss degradation, 0.65 times the rank of fixed-CCR proxy degradation, and 0.05 times directional support. The proxy fuses the raw quantum mass with the other branch's raw Poss mass. Final candidate pairs are instead compared after reliability discounting, using the mean of validation accuracy and macro-F1; fewer blocks break ties. The parameters of the finally selected structure are fitted by the evidential objective.

B.3 Recorded training variants and evaluation metrics

Protocol version. The three-term QEGN formulation is given completely by Eqs. (13)–(17) and is used in the generator comparison described in Appendix B.5. This section records the distinct recipes used by the relation and fusion experiments in Section 5. Four-class relation heads additionally used a focal-sample likelihood, class-order ranking, and augmented contour control; the compact ResNet-9 CIFAR-10 study used class-order ranking and augmented contour control; multimodal heads used class-order ranking, an empty-mass penalty, and FB error ranking. These are experiment-specific variants, not hidden terms in the three-term objective.

B.3.1 Normalized contour and class order

Put $e(x) = \|\bar{p}(x) - \bar{\pi}(m(x))\|_\infty$. Let $\mathrm{Tail}_{.2}$ denote the mean over the largest 20% of values in a batch. The four-class contour loss is

$$\mathcal{L}_c = \mathbb{E}_x \frac{1}{n} \|\bar{p}(x) - \bar{\pi}(m(x))\|_2^2 + \tfrac{1}{2}\,\mathrm{Tail}_{.2}(e(x)^2). \tag{62}$$

The four-class augmented term is $\nu(\mathcal{L}_c - \tau_c) + \frac{\rho}{2}[\mathcal{L}_c - \tau_c]_+^2$, with $\tau_c = 0.005$, $\rho = 60$, initial $\nu = 0.5$, and the epochwise update $\nu \leftarrow [\nu + 8(\mathcal{L}_c - \tau_c)]_+$. This training tolerance applies to the composite loss. CIFAR-10 uses plain normalized-contour mean squared error, with the same augmented-penalty form and the coefficients in Appendix B.2. The separately reported P95 statistic is the 95th percentile of $e(x)$.

The order loss averages

$$\mathrm{softplus}\left(-\,\mathrm{sign}(\bar{p}_i - \bar{p}_j)\,\frac{\bar{\pi}_i - \bar{\pi}_j}{0.05}\right) \tag{63}$$

over pairs with $|\bar{p}_i - \bar{p}_j| > 0.02$. Multimodal learning uses the tolerance form

$$\mathcal{L}_c = \mathbb{E}[e(x) - \tau]_+^2 + \tfrac{1}{2}\,\mathrm{Tail}_{.2}([e(x) - \tau]_+^2) + 0.05\mathbb{E}\frac{1}{n}\|\bar{p} - \bar{\pi}\|_2^2, \tag{64}$$

where $\tau = 0.05$ for MSLCC and 0.06 for CREMA-D. Thus the tolerance and all reported contour errors use normalized contours.

B.3.2 Four-class relation observations and direction ranking

The R observations are produced by two independently fitted groups of 128 bootstrap linear classifiers on frozen features, one for training observations and one for evaluation observations. Both groups use only the classifier-fitting labels. For each input and channel, the 128 sigmoid scores are ranked; the largest $\mathrm{round}(128 p_i)$ scores receive membership one and the remainder zero. A shared ensemble index across channels then defines one focal-set observation. This construction approximates the input marginal to the rounding resolution while retaining cross-channel response dependence. The sparse likelihood is the average negative log mass of these observed focal masks, including empty masks.

Let N_{ji} count true class j among predictions of class i in the calibration role. The reverse confusion estimate uses a uniform prior:

$$r_{ij} = \frac{N_{ji} + 0.5/n}{\sum_k N_{ki} + 0.5}. \tag{65}$$

Its off-diagonal normalization gives d_{ij} in Eq. (15). Row reliability is $w_i = [N_i/(N_i+30)][E_i/(E_i+5)]$, where N_i is prediction support and E_i the number of errors. The shuffled control permutes off-diagonal directions within each row while retaining N_i, E_i, and total error strength.

The output mass is conditioned on nonempty focal sets for the directional statistic. Let π^+ and t_{ij}^+ be its contour and centered pair moment. The standardized pair relation is

$$\gamma_{ij}(x) = \frac{t_{ij}^+(x)}{\sqrt{\pi_i^+(1-\pi_i^+)\pi_j^+(1-\pi_j^+) + 10^{-5}}}. \tag{66}$$

The sample target is $s_{ij}(x) = p_i(x) r_{ij} + p_j(x) r_{ji}$. Weighted pairwise softplus ranking at temperature 0.10 aligns the order of the six γ_{ij} values with the order of s_{ij}; target differences below 0.002 are

omitted, and weights are capped absolute target differences at 0.25. For anchor i, the aggregate relation to j is

$$G_{ij} = \frac{\sum_{x:\widehat{Y}(x)=i} h_{ij}(x)\gamma_{ij}(x)}{\sum_{x:\widehat{Y}(x)=i} h_{ij}(x)}, \qquad h_{ij} = \frac{2\pi_i^+\pi_j^+}{\pi_i^+ + \pi_j^+}. \tag{67}$$

A second ranking term aligns each G_i with d_i, weighted by w_i and the absolute directional gap. $\mathcal{L}_d$ averages the sample and aggregate ranking terms with equal weight.

Held-out direction Spearman compares the 12 off-diagonal G_{ij} entries with the directions recomputed from held-out labels and frozen predictions. Top-alternative recovery is

$$\frac{1}{n}\sum_i \mathbf{1}\left\{\arg\max_{j\neq i} G_{ij} = \arg\max_{j\neq i} d_{ij}^{\text{heldout}}\right\}. \tag{68}$$

The reported six-task mean first averages the three seeds for each task and then averages the tasks equally. Conditional mass TV compares m^+ with the nonempty-conditioned Poss distribution formed from the exact raw output contour.

B.3.3 CIFAR-10 PAIR GUIDANCE

For each frozen feature vector, 32 independent feature-dropout masks at rate 0.15 are applied with inverted-dropout scaling. The fixed classifier produces perturbed scores, which are rank-coupled per class to the input contour in the same manner as above, now at a resolution of 32 samples. These memberships supply centered, variance-normalized targets for all 45 class pairs. A pair is included when the product of empirical Bernoulli variances is at least 10^{-3}. The loss is a masked mean squared error on these targets; this experiment uses no focal-sample likelihood term. For evaluation, the classifier's row-normalized confusion matrix is symmetrized, its diagonal removed, and its 45 unique pairs are compared with test-mean exact pair masses and commonalities.

B.3.4 MULTIMODAL RESCUE RELATIONS AND UNCERTAINTY RANKING

For anchor i and alternative j, define the evidence score

$$R_{ij}(m) = \sum_{A\supseteq\{i,j\}} \frac{m(A)}{|A|-1}, \qquad i \neq j. \tag{69}$$

A composite focal mass is distributed across its possible alternatives to the anchor. A row-normalized rescue matrix counts errors of the current branch that are correctly classified by the other modality. On such samples, a relation cross-entropy at temperature 0.10 favors the true alternative in the anchor's R row. A second term matches the mean row for each anchor to the global rescue direction. Their weights are 0.65 and 0.35. Rows without observed recoverable errors use that branch's error counts to define the direction.

The multimodal objective includes an additional FB error-ranking term,

$$\mathcal{L}_u = \text{softplus}\left[\frac{\mathbb{E}(H_{\text{FB}} \mid \text{backbone correct}) - \mathbb{E}(H_{\text{FB}} \mid \text{backbone incorrect})}{0.10}\right], \tag{70}$$

computed for batches containing both groups. This recorded auxiliary criterion gave entropy an explicit prediction-error ordering objective, in addition to the maximizing FB preference. It is not included in the revised three-term QEGN objective. Its recorded coefficients and the other settings are listed below.

Setting	MSLCC	CREMA-D
Contour / class order	25 / 0.10	24 / 0.10
Relation direction / empty mass	1.25 / 0.25	1.50 / 0.30
FB weight, start–end	0.005–0.035	0.005–0.040
FB error-ranking weight	0.05	0.08
Head batch size	128	64
Head learning rate	0.003	0.003

B.4 Evidence discounting and learnable CCR

Let $h = H_{\mathrm{FB}}/\log(2^n - 1)$. Source reliability is fitted on the fusion role as

$$\alpha(h, \widehat{y}) = \sigma(b_{\widehat{y}} - \mathrm{softplus}(v)h), \qquad m^{\alpha}(A) = \alpha m(A) + (1-\alpha)\mathbf{1}\{A = \Omega\}. \tag{71}$$

The parameters minimize binary cross-entropy against source correctness. Native Poss and quantum masses have separately fitted reliability maps; IsoPoss uses the quantum reliability map and an independent mass made from the normalized quantum output contour. Thus each listed CCR method fuses discounted evidence in its stated representation.

For two source sets A, B and output set F, a source-pattern-controlled target register has conditional mass

$$K_g(F \mid A, B; x) = \prod_{i \in F} g_{i, z_i(A), z_i(B)}(x) \prod_{i \notin F} [1 - g_{i, z_i(A), z_i(B)}(x)], \tag{72}$$

and the fused mass is

$$m_g(F \mid x) = \sum_{A,B} m_1^{\alpha}(A \mid x) m_2^{\alpha}(B \mid x) K_g(F \mid A, B; x). \tag{73}$$

Conditioned on the source membership pattern, the target R_y angle is $2 \arcsin \sqrt{g}$; tracing out the source registers gives this exact target distribution. Fixed CCR follows from $g_{00} = g_{10} = g_{01} = 0$ and $g_{11} = 1$. The learned kernel uses

$$g_{i,10} = \sigma(b_{i,10} + \mathrm{softplus}(v_{i,10})(h_2 - h_1)), \tag{74}$$

$$g_{i,01} = \sigma(b_{i,01} + \mathrm{softplus}(v_{i,01})(h_1 - h_2)), \tag{75}$$

$$g_{i,11} = \sigma(b_{i,11}), \qquad g_{i,00} = 0. \tag{76}$$

Source entropy differences thus modulate retention under single-source support. The initial retention is 0.15 for patterns 10/01 and 0.98 for 11.

Fusion parameters minimize pignistic NLL with sample weight $1 + 1.5\mathbf{1}_{\mathrm{disagree}} + 2\mathbf{1}_{\mathrm{one\ correct}} + 0.5\mathbf{1}_{\mathrm{both\ correct}}$, together with an empty-mass penalty and a penalty on single-source retention. Empty penalties are selected from $\{0.01, 0.03, 0.08\}$ and conjunction penalties from $\{0, 0.002\}$. Learning rate is 0.02; maximum epochs and patience are 120/20 for MSLCC and 160/24 for CREMA-D. Batch sizes are 512 and 48. Checkpoints and fixed-versus-learned fusion choices are selected by validation macro-F1. QEGN-selected uses learned CCR for all Munich seeds, fixed/learned/learned for Berlin, and learned/learned/fixed for CREMA-D.

Validation-selected residual fusion. The retained residual variant combines the selected fusion mass with the singleton BPA of source 2:

$$m_{\mathrm{residual}}(A) = (1-\alpha) m_{\mathrm{U2}}^{\mathrm{singleton}}(A) + \alpha m_{\mathrm{selected}}(A). \tag{77}$$

Here $m_{\mathrm{U2}}^{\mathrm{singleton}}(\{\omega_i\})$ is the source-2 class probability and is zero on composite focal sets. Source 2 is optical in MSLCC and video in CREMA-D. The fixed-versus-learned CCR choice in m_{selected} uses validation macro-F1. The global mixing coefficient α is selected separately by validation accuracy, with macro-F1 and then α breaking ties. For seeds 0/1/2, the selected coefficients are 0.4/0.3/0.2 (Munich), 0.6/0.8/0.35 (Berlin), and 1.0/0.9/1.0 (CREMA-D).

Resolvable disagreement is the subset where the two unimodal predictions differ and exactly one matches the acted or scene label. Its recovery score is fused accuracy on that subset. The three-seed subset sizes are 7,652/7,412/7,060 for Munich, 6,430/6,035/6,082 for Berlin, and 242/249/255 for CREMA-D. Human-conflict evaluation instead selects CREMA-D clips with different untied voice and face majority judgments and an untied audiovisual majority judgment. The targets are audiovisual vote frequencies, with subset sizes 241/252/236. NLL is cross-entropy against that empirical distribution; Brier is the sample-averaged sum of squared class-probability differences.

B.5 Uncertainty evaluation and circuit-generation settings

Scope and common inputs. This controlled study keeps one trained MVRSD-4/ResNet-34 backbone fixed and compares circuit generators across three quantum seeds. Its fitting, relation-calibration,

validation, threshold-calibration, and evaluation roles contain 12,334, 1,542, 772, 773, and 3,513 images, respectively. All methods reuse the same logits, eight-dimensional class activations, temperature 2.7623567, activation standardization, and relation target. The latter is estimated from off-diagonal errors on the relation role. The evaluation role contains 335 errors of the common classifier; local-alternative metrics use these same images. Three quantum seeds do not represent three independently retrained backbones.

The 24 candidates and their semantics. Table 8 enumerates the shared library using one-based hypothesis indices. For each listed support S, every listed t instantiates one distinct $B_{S\to t}$ from Eq. (46). All pairs precede all triples; supports and then targets are lexicographically ordered. Each block applies a CNOT parity map onto t, an R_z rotation, an R_y rotation, and the inverse parity map. It therefore instantiates the same fixed focal-pair revision rule in Eq. (47), with context sign $(-1)^{|D\cap(S\setminus\{t\})|}$. Pair and triple instances cost two and four logical CNOTs, respectively. The terminology refers to support size, not to a claim that the block exclusively changes correlations of that order.

Table 8: The identical 24-module candidate library used by all three joint generators. Each target listed in a row defines a separate module; controls are $S \setminus \{t\}$. Labels translate implementation indices $0, 1, 2, 3$ to hypotheses $1, 2, 3, 4$.

Pair support S	Targets t	Count	Triple support S	Targets t	Count
$\{1,2\}$	$1,2$	2	$\{1,2,3\}$	$1,2,3$	3
$\{1,3\}$	$1,3$	2	$\{1,2,4\}$	$1,2,4$	3
$\{1,4\}$	$1,4$	2	$\{1,3,4\}$	$1,3,4$	3
$\{2,3\}$	$2,3$	2	$\{2,3,4\}$	$2,3,4$	3
$\{2,4\}$	$2,4$	2			
$\{3,4\}$	$3,4$	2			
Total		12	Total		12

Three adapted search mechanisms. We adapt recent search mechanisms to this evidential library and objective, rather than reproduce their original application-specific models. With one architecture logit η_ℓ per candidate, put $q_\ell = \sigma(\eta_\ell)$. *Sampled-DQAS*, motivated by sampled differentiable search (Chen et al., 2026a; Zhang et al., 2022), draws four independent binary inclusion vectors per minibatch. Gate parameters differentiate through the sampled circuits; a score-function estimator updates $\boldsymbol{\eta}$, using the mean loss of the other three paths as a leave-one-out baseline. Its search objective is the expected evidential loss of individual circuits. *Gumbel* adapts differentiable gate selection (Afane et al., 2026). A hard binary Gumbel draw multiplies both angles of each block, so its forward pass is an executable unitary circuit; the backward pass uses the soft angle-gating surrogate. Temperature decreases linearly from 1 to 0.2. Gate parameters update per minibatch, while architecture gradients are averaged and applied once per epoch. *RhoDARTS-core* adapts density-matrix search (Kumar et al., 2025): at candidate ℓ, the search state evolves as

$$\rho \longmapsto (1 - q_\ell)\rho + q_\ell B_\ell \rho B_\ell^\dagger. \tag{78}$$

This is a completely positive trace-preserving mixture of skip and revision; architecture and gate parameters update per minibatch. No application-specific chemistry, adversarial, or noise-training objective is introduced.

Shared fitting and hard export. All search logits start at zero. For each quantum seed, all methods share one full-library parameter initialization. Search runs for eight epochs and fixed-structure fitting for 30 epochs, with batch size 512, AdamW learning rate 0.003, weight decay 10^{-4}, and gate-parameter gradient clipping at one. Architecture logits use Adam at 0.01 without weight decay. The objective is exactly $10\mathcal{L}_{\text{contour}} + \mathcal{L}_{\text{relation}} - \lambda_f \mathcal{H}_{\text{FB}}$, with $\lambda_f = 0.03$ during search and 0.10 during final fitting. There is no architecture-entropy or gate-count penalty.

At every search epoch, the modern generators export 12 distinct modules: first retain the highest-scoring module for each target, then fill remaining places by score, preserving library order. The exported head retains four input and four adjustment R_y gates and has 740 formal trainable parameters.

Validation checkpoint selection prioritizes normalized-contour P95 at most 0.10, then lower AURC; when no checkpoint meets the tolerance, it prioritizes lower contour P95. Validation uses the exported hard circuit, including for density-matrix search.

The three generators inherit the exported search parameters before fixed-structure fitting, using a new optimizer.

Hard-circuit evaluation. RhoDARTS-core evaluates the loss of an architecture-averaged mass, whereas sampled search averages per-circuit losses. Max normalization, alternative normalization, and FB entropy make these objectives distinct. Every reported generator comparison uses the exported deterministic 12-module pure-state circuit before optional refinement.

An equivalent batched PyTorch executor is checked against the retained PennyLane implementation; it changes execution, not the circuit or objective. Computation uses an NVIDIA GeForce RTX 5070 Ti Laptop GPU, PyTorch 2.13.0+cu130, and PennyLane 0.45.1. The reported uncertainty score is FB entropy, while selective decisions remain those of the frozen classifier. AURC is the mean cumulative error after sorting images by increasing uncertainty; lower is better. Results use three quantum seeds, and error bars denote sample standard deviation. Comparisons with separately trained evidential classifiers report their native decisions and therefore assess complete selective-prediction systems.

Classical reference systems. The MVRSD-4 CE backbone uses AdamW for 30 epochs, learning rate 3×10^{-4}, weight decay 10^{-4}, and batch size 48. EDL (Sensoy et al., 2018) uses Adam at 0.001 for 50 epochs, with KL annealing over ten epochs. DAEDL (Yoon & Kim, 2024) uses spectral normalization, feature dropout 0.5, exponential evidence, fixed KL coefficient 0.05, and Adam at 0.001 with per-epoch decay 0.95. Its class-conditional Gaussian densities and density normalization are fitted on the task-fitting role. These baselines use their respective evidence objectives and validation checkpoint selection on the shared backbone architecture. The DAEDL implementation uses that common backbone rather than reproducing the authors' original architecture. Each classical reference uses one training seed, and each quantum generator uses three seeds with the same frozen CE backbone.

Evaluation metrics. QEGN uses normalized nonempty FB entropy as its uncertainty score and retains the frozen CE classifier's decisions; EDL and DAEDL use their native decisions. For N examples sorted by increasing uncertainty, retained-prefix risk and AURC are

$$r_k = \frac{1}{k}\sum_{\ell=1}^{k} \mathbf{1}\{\widehat{y}_{(\ell)} \neq y_{(\ell)}\}, \qquad \mathrm{AURC} = \frac{1}{N}\sum_{k=1}^{N} r_k. \tag{79}$$

Equal uncertainty scores use the expected error within their tied group. Acc90 is accuracy at 90% coverage; error AUROC measures discrimination between correct and incorrect predictions. Local-alternative evaluation uses the same 335 erroneous CE predictions for every method. It excludes the erroneous anchor class, renormalizes each alternative distribution, and evaluates the probability and rank of the true class. For the QEGN head, the alternative distribution is defined by Eq. (15); classical references use their class probabilities with the same anchor excluded. Alternative NLL is the mean negative log true-class probability, and alternative Top-1 is its recovery as the most likely remaining class.