QUANTUM-ASSISTED PROTOTYPE-GUIDED REFINEMENT FOR ROBUST MEDICAL IMAGE CLASSIFICATION UNDER LABEL NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Labels of medical images are often corrupted by annotation errors, diagnostic ambiguity, and uncertainty in report-derived findings, and deep networks trained on such labels memorize the errors. Prototype-based label correction can repair corrupted supervision, but only in an embedding where classes are separated and concentrated; otherwise erroneous assignments contaminate the prototypes and reinforce the errors they are meant to correct. We propose **HQProNet**, which couples trusted-anchor prototype refinement with a compact variational quantum circuit (VQC) that defines the matching geometry. Prototypes are initialized from a small audited anchor set, temperature-scaled cosine similarities yield refined soft targets, and confidence-gated moving-average updates adapt the prototypes during training. The VQC maps a unit-norm 256-dimensional amplitude vector to eight Pauli-Z expectations, each a trainable quadratic form with eigenvalues in $[-1, 1]$, using 120 rotation angles. A conditional analysis shows that prototype matching recovers the clean label when the number of trusted anchors, the within-class covariance, and the cosine margin jointly satisfy an explicit condition. On six MedMNISTv2 datasets under symmetric and pairwise corruption, HQProNet attains the best macro-F1 in 34 of 36 conditions, and on CheXpert it is best on all six metrics under two uncertainty policies. With the prototype mechanism fixed, replacing the VQC by linear, nonlinear, or parameter-matched classical projectors lowers macro-F1 by at least 2.2 points in every setting and reduces the cosine margin and refinement accuracy: prototype refinement is the primary source of robustness, and the VQC supplies the most favorable geometry among the projectors tested.

1 INTRODUCTION

Medical image labels are frequently imperfect: annotators err, findings are ambiguous, and labels mined from radiology reports carry explicit uncertainty (Karimi et al., 2020; Frénay & Verleysen, 2014; Song et al., 2023). The corruption is often structured or class dependent (Xue et al., 2022), and deep networks trained on it memorize the errors, which degrades generalization and calibration (Englesson & Azizpour, 2024; Ju et al., 2022; Shu et al., 2019). Loss correction, sample reweighting, small-loss selection, semi-supervised separation, and iterative relabeling mitigate the problem (Shu et al., 2019; Han et al., 2018; Li et al., 2020; Aliev et al., 2018; Patrini et al., 2017; Arazo et al., 2019; Tanaka et al., 2018; Yi & Wu, 2019; Northcutt et al., 2021). Prototype-based correction is particularly direct: it replaces an observed label with the class whose prototype is most similar to the sample (Zhang et al., 2025; Snell et al., 2017). Its reliability, however, is a property of the embedding. If classes are not separated in angle or are widely dispersed, a corrupted sample is matched to the wrong prototype, the wrong assignment contaminates that prototype, and the error is reinforced. This dependence raises a concrete design question: which projector gives prototype correction a favorable geometry at small parameter cost? Variational quantum circuits (VQCs) have been used in medical-image pipelines as compact classification heads (Landman et al., 2022; Liu et al., 2021; Khatoniar et al., 2024; Mazher et al., 2024; Yousif et al., 2024; Ajlouni et al., 2023), but almost always under clean supervision, and their measured feature space has not been examined as a geometry for label correction. A VQC applied to an amplitude-encoded unit vector has a specific structure: each Pauli-Z readout is a quadratic form of the input whose matrix is generated by

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.

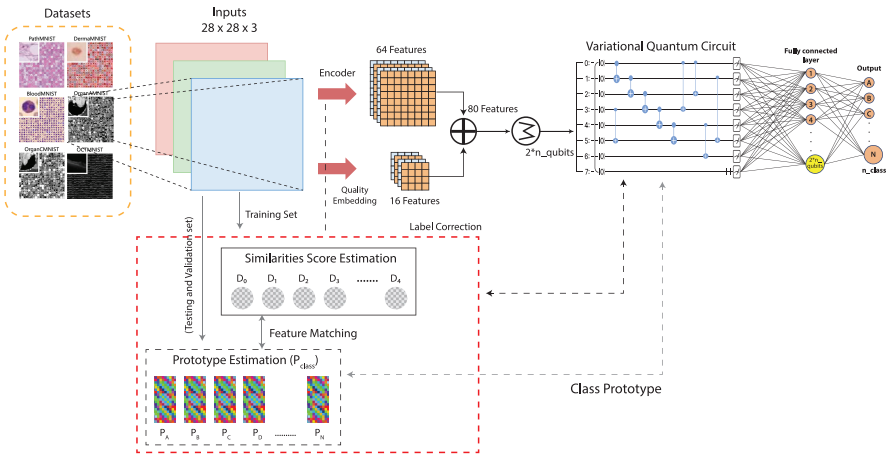


Figure 1: **Overview of HQProNet.** A classical encoder and a quality branch produce an 80-dimensional representation, which is projected to a unit-norm 256-dimensional vector and amplitude encoded into eight qubits (2^Q with $Q = 8$). The VQC returns eight Pauli-Z expectations that define the feature space of both the classifier and the prototype module. A small audited anchor set initializes one prototype per class; temperature-scaled cosine similarities to the prototypes form refined soft targets, and high-confidence assignments update the prototypes by an exponential moving average. The classifier is trained on a weighted combination of observed and refined targets; refined targets and prototype updates are detached from gradient computation. The prototypes P_k and similarity scores D_k in the figure correspond to $\mathbf{p}_{c,e}$ and $s_{c,e}$ (Eqs. (5)–(6)) with $c = k$.

a shared unitary and has eigenvalues in $[-1, 1]$ (Sec. 3). The readouts are therefore bounded and invariant to the input scale, and the entire projection is controlled by a few rotation angles. These properties match the needs of cosine prototype matching, which uses only directions and must be estimated from few trusted samples.

We propose the Hybrid Quantum Prototype Network (**HQProNet**), which uses an eight-qubit, five-layer VQC as the projector of a trusted-anchor prototype-refinement system (Fig. 1). A small audited subset initializes one prototype per class; cosine similarities to the prototypes produce refined soft targets that are mixed with the observed labels; and confidence-gated exponential moving averages update the prototypes as the representation evolves. We evaluate the system under symmetric and pairwise corruption on six MedMNISTv2 datasets and under two report-uncertainty policies on CheXpert, and we attribute its gains through controls that hold the prototype mechanism fixed while replacing the projector. Our contributions are: **(i)** a prototype-refinement method whose matching geometry is defined by the measured features of a VQC, with an explicit training procedure for multi-class and multi-label targets; **(ii)** a conditional analysis that links correct prototype matching to the anchor count, within-class covariance, and cosine margin, and links prototype error to the EMA rate and representation drift; **(iii)** a controlled attribution showing that prototype refinement is the primary source of robustness and that the VQC improves margin, concentration, and refinement accuracy over linear, nonlinear, and approximately parameter-matched classical projectors; and **(iv)** circuit, trusted-anchor, geometry, correction, and hyperparameter studies that delimit these gains.

2 Related Work

Hybrid quantum–classical medical imaging. Parameterized quantum circuits have been added to classical pipelines as trainable transformations or classifier heads (Landman et al., 2022; Khatoniar et al., 2024; Liu et al., 2021; Liang et al., 2021; Schuld et al., 2020; Cerezo et al., 2021), and QCNN-style and quantum-transfer models have been evaluated on radiographs, retinal images, and lightweight medical benchmarks (Cong et al., 2019; Dhara et al., 2024; Bokhan et al., 2022; Mazher et al., 2024; Yousif et al., 2024; Ajlouni et al., 2023). These studies assume clean labels. NQNN adapts a quantum classifier to noisy labels (Rahman & Zhuang, 2025), and QAPNet targets corrupted

inputs with patchwise quantum attention (Rahman & Zhuang, 2026); neither uses the measured quantum features to correct labels.

Learning with noisy medical labels. Classical approaches include uncertainty estimation, sample reweighting, small-loss exchange, semi-supervised separation, and iterative cleaning (Karimi et al., 2020; Ju et al., 2022; Han et al., 2018; Li et al., 2020; Ying et al., 2023; Li et al., 2023; Yi et al., 2022; Feng et al., 2023). Methods that exploit a small trusted subset reweight samples or purify labels with it (Hendrycks et al., 2018; Ren et al., 2018; Tu et al., 2023), and BoMD addresses noisy multi-label chest-radiograph classification (Chen et al., 2023). Prototype- and similarity-based methods revise supervision from class-level structure (Algan & Ulusoy, 2020; Yao et al., 2017; Zhang et al., 2025; Snell et al., 2017; Liang et al., 2022; Khosla et al., 2020); their reliability depends on the separation and concentration of the matching embedding, which is the quantity our projector study varies.

Quantum feature maps. Following the feature-map view of quantum models (Schuld & Killoran, 2019; Havlíček et al., 2019), we treat the VQC as the map that defines the geometry of a trusted-anchor refinement system and compare it with classical projectors inside the same system.

3 Proposed HQProNet Framework

HQProNet combines classical feature extraction, an eight-qubit VQC measurement map, and trusted-anchor prototype refinement (Fig. 1). The circuit has five layers of trainable single-qubit rotations with ring-CNOT entanglement and per-qubit Pauli-Z readout (Fig. 4, App. D).

Hybrid feature extraction and quantum embedding. Let $\mathbf{X} \in \mathbb{R}^{H\times W_{\mathrm{img}}\times C_{\mathrm{in}}}$ denote an input image, where H, W_{img}, and C_{in} are its height, width, and channel count. A classical encoder g_{enc} and a quality branch g_{qual}, adapted from the quality-embedding network of Yao et al. (2017), produce $\mathbf{h}_{\mathrm{img}}(\mathbf{X}) \in \mathbb{R}^{64}$ and $\mathbf{h}_{\mathrm{qual}}(\mathbf{X}) \in \mathbb{R}^{16}$; both are trained jointly with the rest of the model. Their concatenation is

$$\mathbf{h}(\mathbf{X}) = [\mathbf{h}_{\mathrm{img}}(\mathbf{X}); \mathbf{h}_{\mathrm{qual}}(\mathbf{X})] \in \mathbb{R}^{80}. \tag{1}$$

Let $Q = 8$ denote the number of qubits and measured Pauli-Z coordinates. A trainable matrix $\mathbf{W}_{\mathrm{amp}} \in \mathbb{R}^{2^Q\times 80}$ produces the pre-encoding vector $\mathbf{v}(\mathbf{X}) = \mathbf{W}_{\mathrm{amp}}\mathbf{h}(\mathbf{X})$, whose unit-norm amplitude vector is

$$\mathbf{z}(\mathbf{X}) = \frac{\mathbf{v}(\mathbf{X})}{\|\mathbf{v}(\mathbf{X})\|_2}. \tag{2}$$

A zero pre-encoding vector is mapped to a fixed computational-basis state. Otherwise, $\mathbf{z}(\mathbf{X})$ is amplitude encoded as $|\psi(\mathbf{X})\rangle = \sum_{b=0}^{2^Q-1} z_b(\mathbf{X})|b\rangle$, where b indexes computational-basis states. A five-layer circuit applies trainable single-qubit rotations and ring-CNOT entanglement. At epoch e, its Pauli-Z expectations form the raw readout $\mathbf{r}_e(\mathbf{X}) \in [-1, 1]^Q$. With a fixed numerical stabilizer $\xi_0 > 0$, the normalized feature used by the classifier and the prototype module is

$$\boldsymbol{\phi}_e(\mathbf{X}) = \frac{\mathbf{r}_e(\mathbf{X})}{\max\{\|\mathbf{r}_e(\mathbf{X})\|_2, \xi_0\}}. \tag{3}$$

Measured geometry. Let $U(\boldsymbol{\theta}_e)$ be the circuit unitary at epoch e and Z_j the Pauli-Z operator on qubit j. Because $\mathbf{z}(\mathbf{X})$ is real and unit-norm, each readout is the quadratic form

$$r_{e,j}(\mathbf{X}) = \mathbf{z}(\mathbf{X})^\top \mathbf{M}_{e,j}\, \mathbf{z}(\mathbf{X}), \qquad \mathbf{M}_{e,j} = \mathrm{Re}\big(U(\boldsymbol{\theta}_e)^\dagger Z_j\, U(\boldsymbol{\theta}_e)\big), \tag{4}$$

where $\mathbf{M}_{e,j}$ is symmetric with eigenvalues in $[-1, 1]$. The VQC thus computes Q bounded quadratic forms of the amplitude vector, all generated by one unitary with $3QL = 120$ rotation angles for $L = 5$ layers. A linear projector instead computes linear forms of $\mathbf{z}(\mathbf{X})$, and an MLP applies unconstrained pointwise nonlinearities; Sec. 5 compares the three inside the same refinement system.

Trusted-anchor prototype refinement. Let J be the number of classes, $c \in \{1, \ldots, J\}$ a class index, $\mathcal{A}_c$ the trusted anchor set for class c, and $m_c = |\mathcal{A}_c|$ its size. The initial prototype is

$$\mathbf{p}_{c,0} = \frac{1}{m_c} \sum_{\mathbf{X}\in\mathcal{A}_c} \boldsymbol{\phi}_0(\mathbf{X}). \tag{5}$$

At epoch e, the stabilized cosine score is

$$s_{c,e}(\mathbf{X}) = \frac{\boldsymbol{\phi}_e(\mathbf{X})^\top \mathbf{p}_{c,e}}{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \|\mathbf{p}_{c,e}\|_2 + \xi_0}. \tag{6}$$

For temperature $T > 0$, the refined soft target is

$$t_{c,e}(\mathbf{X}) = \frac{\exp(s_{c,e}(\mathbf{X})/T)}{\sum_{j=1}^{J} \exp(s_{j,e}(\mathbf{X})/T)}, \tag{7}$$

where j is a competing class index. The hard refined assignment and confidence are $\widehat{y}_e(\mathbf{X}) = \arg\max_c t_{c,e}(\mathbf{X})$ and $\alpha_e(\mathbf{X}) = \max_c t_{c,e}(\mathbf{X})$. For confidence threshold τ, define $\mathcal{G}_{c,e} = \{\mathbf{X} : \widehat{y}_e(\mathbf{X}) = c, \alpha_e(\mathbf{X}) \geq \tau\}$. If $\mathcal{G}_{c,e}$ is nonempty, its mean feature is $\overline{\boldsymbol{\phi}}_{c,e} = |\mathcal{G}_{c,e}|^{-1} \sum_{\mathbf{X} \in \mathcal{G}_{c,e}} \boldsymbol{\phi}_e(\mathbf{X})$ and the prototype is updated as

$$\mathbf{p}_{c,e+1} = (1-\beta)\mathbf{p}_{c,e} + \beta \overline{\boldsymbol{\phi}}_{c,e}, \tag{8}$$

where $\beta \in (0, 1]$ is the EMA rate; if $\mathcal{G}_{c,e}$ is empty, $\mathbf{p}_{c,e+1} = \mathbf{p}_{c,e}$. Within an epoch, refined targets are computed for each minibatch from the current prototypes $\mathbf{p}_{c,e}$; the sets $\mathcal{G}_{c,e}$ are formed from the assignments and confidences collected over the full epoch, and Eq. (8) is applied once at the end of the epoch. Anchors initialize the prototypes only and contribute no gradients. Refined targets and prototypes are detached from gradient computation (Algorithm 1).

Learning objectives. Let y^{obs} denote the observed multi-class label, $\boldsymbol{\pi}_e(\mathbf{X})$ the classifier probability vector, and $\mathbf{t}_e(\mathbf{X}) = [t_{1,e}(\mathbf{X}), \ldots, t_{J,e}(\mathbf{X})]^\top$ the refined soft target. The multi-class objective is

$$\mathcal{L}_e^{\text{mc}} = (1-\lambda_e)\,\mathrm{CE}\big(\mathrm{onehot}(y^{\text{obs}}), \boldsymbol{\pi}_e(\mathbf{X})\big) + \lambda_e\,\mathrm{CE}(\mathbf{t}_e(\mathbf{X}), \boldsymbol{\pi}_e(\mathbf{X}))\,, \tag{9}$$

where $\lambda_e \in [0, 1]$ is the epoch-dependent weight of refined supervision. For CheXpert, let $\mathbf{q}^{\text{obs}} \in [0, 1]^5$ denote the observed target after the selected uncertainty policy and $\mathbf{u}_e^{\text{ml}}(\mathbf{X}) \in [0, 1]^5$ the sigmoid classifier output. For observation o, certain-negative and certain-positive anchors initialize prototypes $\mathbf{p}_{o,e}^-$ and $\mathbf{p}_{o,e}^+$, with scores $s_{o,e}^{\pm}(\mathbf{X})$ computed as in Eq. (6); uncertain entries are excluded from anchor construction. The refined positive probability is $q_{o,e}^{\text{ref}}(\mathbf{X}) = \exp(s_{o,e}^+/T)/[\exp(s_{o,e}^-/T) + \exp(s_{o,e}^+/T)]$, and the multi-label objective is

$$\mathcal{L}_e^{\text{ml}} = (1-\lambda_e)\,\mathrm{BCE}\big(\mathbf{q}^{\text{obs}}, \mathbf{u}_e^{\text{ml}}(\mathbf{X})\big) + \lambda_e\,\mathrm{BCE}\big(\mathbf{q}_e^{\text{ref}}(\mathbf{X}), \mathbf{u}_e^{\text{ml}}(\mathbf{X})\big)\,. \tag{10}$$

4 Why the Refinement Can Be Reliable

The analysis is conditional on the feature map learned at a fixed epoch and does not imply that every VQC learns a favorable geometry. Assumptions, proofs, the anchor-count corollary, the feature-kernel interpretation, and the multi-label extension are in Apps. B–C.

For class c, define the population prototype $\boldsymbol{\mu}_{c,e} = \mathbb{E}[\boldsymbol{\phi}_e(\mathbf{X}) \mid y^\star = c]$ and covariance $\boldsymbol{\Sigma}_{c,e} = \mathrm{Cov}(\boldsymbol{\phi}_e(\mathbf{X}) \mid y^\star = c)$, where $y^\star$ is the latent clean label. For analysis only, let $\widehat{\boldsymbol{\mu}}_{c,e}^{\text{anc}}$ be the mean of the trusted anchors re-encoded by the fixed epoch-e feature map. Assume the anchors are conditionally independent and correctly labeled and that $\|\boldsymbol{\mu}_{c,e}\|_2 \geq \kappa > 0$. The population cosine margin of a clean class-c image is

$$\Gamma_e(\mathbf{X}) = \frac{\boldsymbol{\phi}_e(\mathbf{X})^\top \boldsymbol{\mu}_{c,e}}{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \|\boldsymbol{\mu}_{c,e}\|_2} - \max_{j \neq c} \frac{\boldsymbol{\phi}_e(\mathbf{X})^\top \boldsymbol{\mu}_{j,e}}{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \|\boldsymbol{\mu}_{j,e}\|_2}. \tag{11}$$

Proposition 1 (Clean-anchor prototype reliability)**.** *For failure probability $\delta \in (0, 1)$, define $\Delta_e = \max_c \|\widehat{\boldsymbol{\mu}}_{c,e}^{\text{anc}} - \boldsymbol{\mu}_{c,e}\|_2$. With probability at least $1 - \delta$,*

$$\Delta_e \leq \max_c \left[\sqrt{\frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_{c,e}) \log(2QJ/\delta)}{m_c}} + \frac{4\sqrt{Q}\log(2QJ/\delta)}{3m_c} \right]. \tag{12}$$

If $\Delta_e < \kappa$ and

$$\Gamma_e(\mathbf{X}) > \frac{4\Delta_e}{\kappa - \Delta_e}, \tag{13}$$

then empirical cosine prototype matching recovers $y^\star$ for $\mathbf{X}$.

The condition shows that anchor count alone is insufficient: reliable refinement also requires concentrated classes and a sufficiently large margin.

Proposition 2 (Target-level incorrect mass)**.** *Let $\eta = \Pr(y^{\mathrm{obs}} \neq y^\star)$ be the observed-label error rate and $\rho_e = \mathbb{E}[t_{y^\star,e}(\mathbf{X})]$ the expected refined-target mass on the clean class. The mixed target has expected incorrect mass*

$$\eta_{\mathrm{eff},e} = (1-\lambda_e)\eta + \lambda_e(1-\rho_e), \tag{14}$$

so $\eta_{\mathrm{eff},e} < \eta$ whenever $\rho_e > 1-\eta$ and $\lambda_e > 0$.

For hard refined assignments, ρ_e is the refinement accuracy. Equation (14) concerns the supervision target rather than the classifier's population risk.

Proposition 3 (EMA stability with representation drift)**.** *Define $\mathcal{E}_{c,e} = \|\mathbf{p}_{c,e} - \boldsymbol{\mu}_{c,e}\|_2$, $\zeta_{c,e} = \|\overline{\boldsymbol{\phi}}_{c,e} - \boldsymbol{\mu}_{c,e}\|_2$, and $\omega_{c,e} = \|\boldsymbol{\mu}_{c,e+1} - \boldsymbol{\mu}_{c,e}\|_2$. Then*

$$\mathcal{E}_{c,e+1} \leq (1-\beta)\mathcal{E}_{c,e} + \beta\zeta_{c,e} + \omega_{c,e}. \tag{15}$$

The threshold τ therefore controls the quality of admitted samples, whereas β trades resistance to contamination against the ability to track a changing representation.

5 Experiment Setup and Results

We evaluate HQProNet on MedMNISTv2 under controlled symmetric and pairwise training-label corruption and on CheXpert under two report-uncertainty policies. Synthetic corruption permits direct measurement of label repair because the clean labels remain available; CheXpert evaluates robustness to policy-dependent ambiguity.

Datasets. We use six MedMNISTv2 benchmarks (Yang et al., 2023) at 28×28: BloodMNIST (D1), PathMNIST (D2), OrganAMNIST (D3), OrganCMNIST (D4), DermaMNIST (D5), and OCTMNIST (D6). We additionally use a 10% stratified patient-level CheXpert subset (Irvin et al., 2019) with Atelectasis, Cardiomegaly, Consolidation, Edema, and Pleural Effusion, preserving patient disjointness and label prevalence.

Noise and uncertainty protocols. MedMNIST corruption is applied only to training labels at rate $\eta \in \{10, 30, 50\}\%$; validation and test labels remain clean. Symmetric corruption replaces a selected label uniformly with an incorrect class, whereas pairwise corruption maps it to a prespecified confusable class. On CheXpert, U-zero maps an uncertain entry to zero, while smoothing assigns it a soft target; these policies are not interpreted as known label-flip rates.

Configuration and trusted anchors. The encoder and quality branch produce the 80-dimensional feature of Eq. (1), which $\mathbf{W}_{\mathrm{amp}}$ projects to the 256-dimensional amplitude vector. The official validation split is partitioned into two disjoint subsets: a stratified trusted set with 5% of each class for prototype initialization, and the remainder for model selection and early stopping (Hendrycks et al., 2018; Ren et al., 2018; Tu et al., 2023). We use $\beta = 0.2$, $\tau = 0.7$, $T = 1.0$, and linearly ramp λ_e to $\lambda_{\max} = 0.5$ by epoch 10.

Training and reporting. All methods use Adam with learning rate 10^{-3}, batch size 128, at most 30 epochs, and early stopping with patience 5. We report means over seeds $\{42, 1337, 2024\}$. Baselines use the shared grid $\{10^{-4}, 3 \times 10^{-4}, 10^{-3}\}$ for learning rate and $\{64, 128\}$ for batch size, selected only on the non-trusted validation subset. Experiments use Python 3.12.12, PyTorch 2.8.0+cu126, PennyLane 0.44.0, and two NVIDIA Tesla T4 GPUs. Standard deviations of the controlled, trusted-budget, geometry, repair, and sensitivity studies are in Tabs. 9–19 (Apps. F–I).

Metrics and baselines. MedMNIST performance is reported by macro-F1; CheXpert by macro-averaged accuracy (ACC), area under the ROC curve (AUC), and F1. Because clean MedMNIST labels are known, we also report refinement accuracy (the fraction of training samples whose refined hard label equals the clean label), repair (the fraction of corrupted labels restored to the clean class), and retention (the fraction of clean labels not changed incorrectly). We compare with QCNN (Liu et al., 2021), QTNet (Dhara et al., 2024), HQNet (Mazher et al., 2024), QCCNet (Yousif et al., 2024), A-HQCNN (Ajlouni et al., 2023), DUE-Net (Ju et al., 2022), RWNet (Díaz & Steele, 2021), Co-teaching (Han et al., 2018), DivideMix (Li et al., 2020), and NQNN (Rahman & Zhuang, 2025). All methods use the same splits and corruption realizations. The projector controls additionally share the

Table 1: **Macro-F1 (%, ↑) on six MedMNISTv2 datasets** under symmetric and pairwise training-label corruption at $\eta \in \{10, 30, 50\}\%$. D1–D6 denote BloodMNIST, PathMNIST, OrganAMNIST, OrganCMNIST, DermaMNIST, and OCTMNIST. Best value per column in bold.

	D1			D2			D3			D4			D5			D6		
Method	10	30	50	10	30	50	10	30	50	10	30	50	10	30	50	10	30	50
Symmetric noise																		
QCNN	57.8	52.3	45.7	59.4	56.9	45.4	57.4	51.3	43.2	56.8	52.5	41.6	53.2	45.5	41.3	24.0	20.9	14.4
QTNet	13.1	12.3	12.8	12.0	10.5	18.0	12.0	11.6	11.8	19.2	18.3	19.4	17.2	14.8	12.7	25.9	19.3	9.8
HQNet	13.5	13.6	10.3	12.6	10.4	10.1	14.0	12.8	13.1	19.6	18.1	18.0	16.1	15.3	12.7	24.8	23.7	23.4
QCCNet	13.6	10.9	11.0	13.2	12.9	11.3	13.7	10.3	12.3	19.5	19.4	19.4	22.4	19.7	16.7	26.3	23.8	22.5
A-HQCNN	22.6	22.5	21.9	25.4	23.8	14.9	14.9	14.6	12.4	13.8	10.7	10.1	13.4	10.8	10.9	27.1	24.8	23.7
DUE-Net	74.0	58.2	48.8	72.2	57.9	47.4	28.5	28.3	23.4	78.2	61.8	47.8	40.8	27.4	23.7	51.3	44.5	39.8
RWNet	71.9	53.1	43.0	68.5	54.1	43.7	43.9	34.3	27.7	75.1	58.5	46.0	22.5	21.9	18.9	49.0	43.4	39.9
DivideMix	82.1	81.9	79.2	83.1	80.1	80.1	78.5	74.7	78.3	75.6	75.6	76.5	74.7	73.8	74.7	72.9	65.7	71.1
Co-teaching	81.8	82.8	78.3	82.1	82.8	78.3	79.2	77.4	76.5	76.1	74.7	75.6	73.8	73.8	74.7	72.9	72.0	72.0
NQNN	83.1	82.8	81.0	83.7	81.9	81.9	78.6	77.4	74.7	76.7	**79.2**	74.7	74.1	72.9	73.8	71.4	70.2	70.2
HQProNet	**85.6**	**85.1**	**83.2**	**86.7**	**85.8**	**84.8**	**81.8**	**80.4**	**79.7**	**80.8**	78.8	**77.7**	**76.8**	**76.4**	**75.2**	**74.8**	**74.4**	**73.0**
Pairwise noise																		
QCNN	55.4	50.7	44.3	56.9	55.2	44.0	55.0	49.8	41.9	54.4	50.9	40.4	51.6	44.1	40.1	23.3	20.3	14.0
QTNet	12.7	11.9	12.4	11.6	10.2	7.8	11.7	11.2	11.4	18.9	18.1	19.1	16.7	14.4	12.4	25.2	18.7	9.5
HQNet	13.1	13.2	10.0	12.2	10.1	9.8	13.5	12.4	12.7	19.3	17.8	17.8	15.6	14.8	12.4	24.1	23.0	22.7
QCCNet	13.2	10.5	10.7	12.8	12.5	11.0	13.3	10.0	11.9	19.2	19.1	19.1	21.7	19.2	16.2	25.5	23.1	21.8
A-HQCNN	21.9	21.8	21.2	24.7	23.1	14.5	14.4	14.2	12.0	13.4	10.4	9.8	13.0	10.5	10.6	26.2	24.1	23.0
DUE-Net	71.0	55.8	47.3	69.2	56.1	46.0	27.6	27.4	22.7	75.0	59.2	46.4	39.6	26.6	23.0	49.8	43.2	38.6
RWNet	68.9	51.5	41.7	65.7	57.2	42.4	42.6	33.3	26.9	72.0	56.1	44.6	21.8	21.3	18.3	47.5	42.1	38.7
DivideMix	78.7	78.5	75.9	79.6	76.8	76.8	75.2	71.6	75.0	72.4	72.4	73.3	71.6	70.7	71.6	69.9	63.0	68.1
Co-teaching	78.4	79.3	75.0	78.7	79.3	75.0	75.9	74.2	73.3	72.9	71.6	72.4	70.7	70.7	71.6	69.9	69.0	69.0
NQNN	79.6	79.3	77.6	80.2	78.5	78.5	75.3	74.2	71.6	73.5	**75.9**	71.6	71.0	69.9	70.7	68.4	67.3	67.3
HQProNet	**82.0**	**81.6**	**79.7**	**83.1**	**82.2**	**81.2**	**78.4**	**77.0**	**76.4**	**77.4**	75.5	**74.4**	**73.6**	**73.2**	**72.1**	**71.7**	**71.3**	**70.0**

encoder, classifier, anchor draw, prototype mechanism, optimizer, stopping rule, and tuning budget (App. E, Tab. 8).

Symmetric and pairwise corruption. HQProNet has the highest macro-F1 in 34 of the 36 dataset–noise conditions of Tab. 1; NQNN is higher on D4 at 30% under both corruption types (79.2 vs. 78.8 and 75.9 vs. 75.5). At 50% corruption, the six-dataset mean of HQProNet is 78.9% under symmetric and 75.6% under pairwise noise. Pairwise corruption is harder because repeated class-to-class flips bias the prototype of the receiving class in a consistent direction.

Quantum baselines. QTNet, HQNet, QCCNet, and A-HQCNN remain within 15 points of the uniform-guessing macro-F1 of $1/J$ (9–25% on these datasets) in every condition of Tab. 1: under the shared protocol of 28×28 inputs, up to 11 classes, corrupted labels, and a 30-epoch budget with the shared tuning grid, they do not learn beyond chance. These models were proposed as classifiers under clean supervision; QCNN learns under the same protocol but stays well below the noisy-label methods.

CheXpert uncertainty policies. HQProNet is best on all six macro metrics in Tab. 2, with 91.5% AUC and 77.2% F1 under smoothing and 81.0% AUC and 70.6% F1 under U-zero. These results measure robustness to two treatments of uncertain report-derived labels, not to a known flip rate. The ACC, AUC, and F1 of the controlled projectors are in Tab. 13.

Table 2: **CheXpert** (%, ↑) under smoothing (Sm) and U-zero (U0), macro-averaged over five observations on the 10% patient-level subset.

	Sm			U0		
Method	ACC	AUC	F1	ACC	AUC	F1
QCNN	61.0	62.7	54.3	59.9	53.2	22.1
QTNet	13.1	12.0	11.1	9.2	17.2	23.8
HQNet	13.5	12.6	12.8	9.6	16.1	22.8
QCCNet	13.6	13.2	12.6	9.5	22.4	24.2
A-HQCNN	22.6	25.4	13.7	13.8	13.4	24.9
DUE-Net	78.2	76.2	26.1	82.6	40.8	47.1
RWNet	75.9	72.3	40.4	79.3	22.5	45.0
DivideMix	86.7	87.7	74.2	79.8	78.9	68.9
Co-teaching	86.4	86.7	74.8	80.3	77.9	68.9
NQNN	87.7	88.4	74.2	81.0	78.2	67.4
HQProNet	**90.3**	**91.5**	**77.2**	**85.3**	**81.0**	**70.6**

Component ablation. Tab. 3 separates the two components. Naïve trains the shared encoder and classifier on observed labels only. Adding the prototype mechanism with a linear projector (Linear+Proto) raises the six-dataset mean at 50% symmetric noise from 40.2 to 75.5, whereas the VQC without prototypes (VQC-only) reaches 66.5; combining both gives 78.9. HQProNet is

Table 3: **Component ablation** (macro-F1, %, ↑). D1–D6: MedMNISTv2 under 10/30/50% symmetric label noise; Chx: CheXpert under smoothing (Sm) and U-zero (U0).

	D1			D2			D3				
Variant	10	30	50	10	30	50	10	30	50		
Naïve	49.7	46.5	41.6	54.6	52.0	46.4	48.0	44.5	38.2		
Linear+Proto	81.9	81.1	80.2	84.7	83.7	81.2	79.3	73.8	74.6		
VQC-only	79.3	74.9	73.0	82.8	79.2	75.6	77.5	71.2	65.8		
HQProNet	**85.6**	**85.1**	**83.2**	**86.7**	**85.8**	**84.8**	**81.8**	**80.4**	**79.7**		
	D4			**D5**			**D6**			**Chx**	
Variant	10	30	50	10	30	50	10	30	50	Sm	U0
Naïve	47.3	43.6	38.0	45.7	42.6	37.9	45.1	39.0	38.8	56.7	54.1
Linear+Proto	73.0	74.6	72.8	70.3	72.2	73.0	72.1	72.9	71.1	74.8	63.9
VQC-only	70.2	65.9	63.9	64.2	62.1	58.8	68.3	66.5	62.1	72.9	65.6
HQProNet	**80.8**	**78.8**	**77.7**	**76.8**	**76.4**	**75.2**	**74.8**	**74.4**	**73.0**	**77.2**	**70.6**

best in every column: refinement repairs the supervision, and the VQC shapes the space in which the repair operates.

Circuit depth and qubit count. The five-layer, eight-qubit circuit is best on all six datasets (Tab. 4), and three layers are worse than five at every qubit count and dataset. Twelve qubits lower macro-F1 at both depths. Because $\mathbf{W}_{\mathrm{amp}} \in \mathbb{R}^{2^Q \times 80}$ grows with Q (1,280, 20,480, and 327,680 parameters for 4, 8, and 12 qubits), this study varies circuit width and classical projection capacity together, and we do not attribute the 12-qubit decrease to the circuit alone.

Table 4: **Circuit ablation** (macro-F1, %, ↑) under 50% symmetric noise. L: variational layers; Q: qubits. Best value per row in bold.

	5 layers			3 layers		
Data	4Q	8Q	12Q	4Q	8Q	12Q
D1	80.6	**83.2**	66.0	77.7	79.4	64.3
D2	82.0	**84.8**	67.0	78.4	80.3	65.1
D3	76.3	**79.7**	62.6	71.4	73.4	61.7
D4	74.6	**77.7**	60.9	70.6	72.5	60.6
D5	71.4	**75.2**	59.8	69.5	70.6	59.5
D6	71.3	**73.0**	58.8	69.6	70.4	57.9
Mean	76.0	**78.9**	62.5	72.9	74.4	61.5

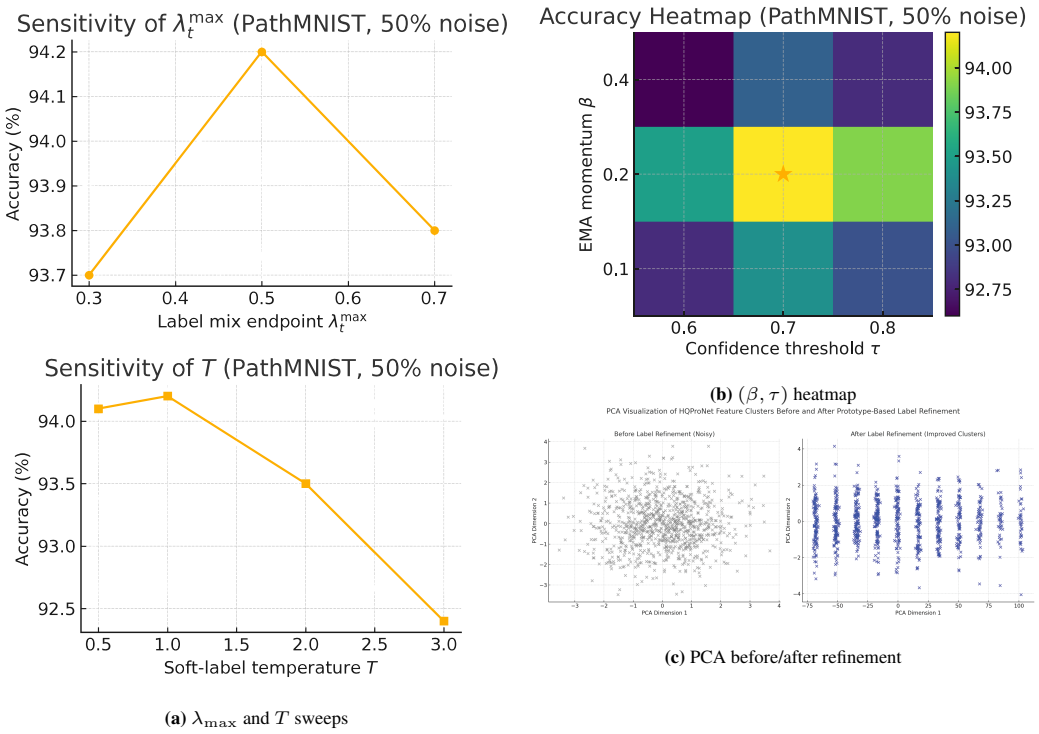
Hyperparameter sensitivity and embedding visualization. Fig. 2 reports a PathMNIST accuracy case study and a qualitative PCA view of the embedding before and after refinement. On the primary metric averaged over all six datasets, the defaults lie on a plateau: $\lambda_{\max} \in \{0.4, 0.6\}$ and $T = 1.5$ stay within 0.3 macro-F1 points of the default, $T = 3.0$ lowers the symmetric/pairwise means from 78.1/74.8 to 76.4/73.2, and $(\beta, \tau) = (0.2, 0.7)$ is the best cell of the EMA–confidence grid (Tabs. 18–19, Fig. 8).

What the VQC contributes. Tab. 5 separates two comparisons. Without prototypes, the VQC raises macro-F1 at 50% symmetric noise from 39.7 (Classical-only) to 66.5 (VQC-only) but stays below every prototype-based variant: the circuit alone does not repair corrupted supervision. With the prototype mechanism, anchor draw, classifier, optimizer, and stopping rule fixed, replacing the VQC by the strongest classical projector (MLP+Proto, 8,488 parameters) lowers macro-F1 by 2.4, 2.4, 2.2, and 2.6 points in the four settings, and replacing it by the approximately parameter-matched Budget-MLP+Proto (110 parameters) lowers it by 3.1, 3.1, 2.9, and 3.4 points. The same substitution by MLP+Proto lowers the mean cosine margin from 0.246 to 0.214, raises the covariance trace from 0.185 to 0.217, and lowers refinement accuracy from 88.0% to 85.2% (Tab. 16), the quantities that enter the recovery condition of Proposition 1. The VQC is therefore

Table 5: **Controlled projector comparison.** Mean macro-F1 (%) and geometry with the pipeline fixed; only the projector and, in the upper block, the prototype mechanism change.

Variant	Par.	S50	P50	Sm	U0	Mar.	Ref.
Without prototype refinement							
Classical-only	0	39.7	38.3	56.1	53.6	0.093	–
VQC-only	120	66.5	63.0	72.9	65.6	0.194	–
With prototype refinement							
Linear+P	2,056	75.5	71.7	74.8	63.9	0.180	83.7
MLP+P	8,488	76.5	73.2	75.0	68.0	0.214	85.2
B-MLP+P	110	75.8	72.5	74.3	67.2	0.201	84.5
HQProNet	**120**	**78.9**	**75.6**	**77.2**	**70.6**	**0.246**	**88.0**

Linear+P, MLP+P, B-MLP+P: Linear, MLP, and Budget-MLP projectors with prototypes. Par.: projector parameters, excluding the $80 \to 256$ projection $\mathbf{W}_{\mathrm{amp}}$ (20,480 parameters) shared by all projector variants. S50/P50: six-dataset MedMNISTv2 mean under 50% symmetric/pairwise noise. Sm/U0: CheXpert under smoothing/U-zero. Mar.: mean cosine margin; Ref.: refinement accuracy (%), both at 50% symmetric noise.



(a) $\lambda_{\max}$ and T sweeps **(b)** (β, τ) heatmap **(c)** PCA before/after refinement

Figure 2: **Sensitivity and embedding visualization.** (a) Refined-target ceiling $\lambda_{\max}$ (labeled $\lambda_t^{\max}$ in the plot) and temperature T; (b) EMA rate β and confidence threshold τ; both report PathMNIST accuracy at 50% symmetric noise. (c) PCA of the embedding before and after refinement, shown qualitatively. Six-dataset macro-F1 sweeps are in Tabs. 18–19 and Fig. 8.

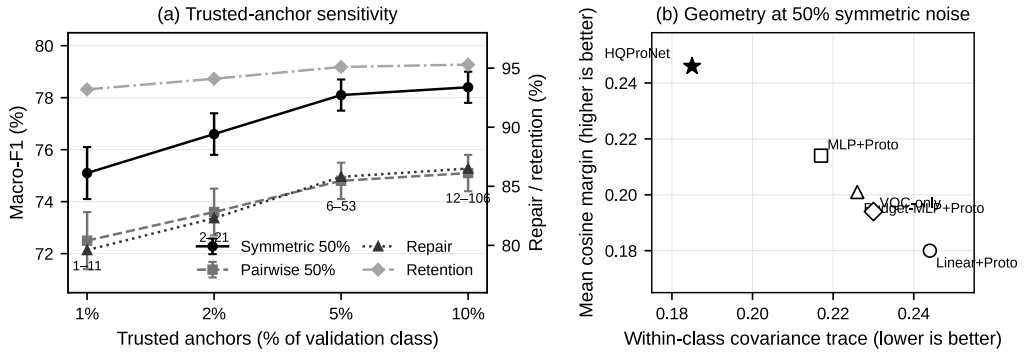


Figure 3: **Trusted-anchor and geometry analyses.** (a) Macro-F1, repair, and retention versus the trusted budget; error bars show standard deviations over three seeds (Tab. 14). Gains are large from 1% to 5% and small from 5% to 10%. (b) Mean cosine margin versus within-class covariance trace at 50% symmetric noise; HQProNet occupies the high-margin, low-dispersion region.

not required for robustness, which prototype refinement provides, but it yields the most favorable geometry for refinement among the projectors tested. Per-dataset results are in App. F (Tabs. 9–13).

Controlled reporting. For injected MedMNIST corruption, repair is the fraction of corrupted labels restored to the latent clean class, and retention is the fraction of initially clean labels not changed incorrectly. Projector controls use identical corruption masks, anchor draws, optimization settings, and prototype rules. They distinguish linear compression (Linear+Proto), generic nonlinear projection (MLP+Proto), and compact classical parameterization (Budget-MLP+Proto). The controlled experiments support a two-stage attribution, not a claim of intrinsic quantum advantage. Prototype refinement is the principal robustness mechanism: with a linear projector, it

raises the six-dataset mean at 50% symmetric corruption from 40.2% to 75.5%. The VQC is not required for this gain, but within the fixed pipeline it gives the highest mean macro-F1 among the projectors tested. Relative to MLP+Proto at 50% symmetric corruption, HQProNet increases the mean cosine margin from 0.214 to 0.246, decreases the within-class covariance trace from 0.217 to 0.185, and raises refinement accuracy from 85.2% to 88.0% (Appendix H). These controls isolate the projector within this architecture; they do not establish superiority over arbitrary classical feature maps. The operational prototype $p_{c,e}$ then evolves through confidence-gated EMA updates; Proposition 3 separates its error into retained initialization error, gated-mean error, and representation drift. The measured margin and covariance are consistent with this mechanism.

Trusted-anchor sensitivity. Most of the benefit of trusted supervision is obtained between 1% and 5% (Tab. 6, Fig. 3): increasing the budget from 1% to 5% raises the 50% symmetric/pairwise means from 75.1/72.5 to 78.1/74.8 and repair from 79.6% to 85.8%, whereas increasing it from 5% to 10% adds 0.3 macro-F1 points under each severe-noise protocol. We therefore use 5% as an empirical operating point rather than a theoretically unique threshold. The trusted-budget and hyperparameter studies are independent reruns in which the default configuration reaches 78.1/74.8, 0.8 points below the corresponding entries of Tabs. 1 and 5; we compare settings only within each study. Per-dataset values are in App. G (Tabs. 14–15, Fig. 6).

Table 6: **Trusted-anchor sensitivity.** The 5% budget is the default.

Budget	Count	S30	S50	P30	P50	Rep.	Ret.
1%	1–11	76.8	75.1	73.8	72.5	79.6	93.2
2%	2–21	78.2	76.6	74.8	73.6	82.3	94.1
5%	**6–53**	**79.3**	**78.1**	**76.0**	**74.8**	**85.8**	**95.1**
10%	12–106	79.7	78.4	76.3	75.1	86.5	95.3

Count: realized trusted samples per class. S30/S50, P30/P50: six-dataset MedMNISTv2 mean macro-F1 (%) under 30%/50% symmetric/pairwise noise. Rep.: corrupted-label repair (%). Ret.: clean-label retention (%).

6 Discussion and Limitations

The controlled results support a specific attribution. Prototype refinement is the main source of robustness: with a linear projector it raises the six-dataset mean at 50% symmetric noise from 40.2 to 75.5. The VQC is not required for this effect, but among the projectors tested it yields the largest cosine margin, the smallest within-class covariance, the most accurate refinement (App. H), and the highest macro-F1 in every controlled setting. Several boundaries remain. HQProNet requires a small trusted anchor set; the anchors are disjoint from the model-selection subset, contribute no gradients, and are shared by all prototype-based controls, but we do not compare with methods that use the same trusted data for label purification (Tu et al., 2023), with the chest-radiograph method BoMD (Chen et al., 2023), or with a clean-label or oracle-correction ceiling. CheXpert smoothing and U-zero are uncertainty policies rather than known corruption rates. The circuit ablation varies the qubit count together with the size of $\mathbf{W}_{\mathrm{amp}}$. All circuits are simulated; finite-shot estimation, device noise, and hardware execution are not evaluated, and the 120 rotation angles indicate compact parameterization rather than a runtime or hardware advantage. Results are means over three seeds without significance testing, and transfer across institutions, scanners, and changing disease prevalence is not established.

7 Conclusion

HQProNet uses eight Pauli-Z expectation values from an eight-qubit VQC to define the cosine-matching geometry for trusted-anchor prototype refinement under corrupted medical-image supervision. Across six MedMNISTv2 datasets, HQProNet achieves the highest macro-F1 in 34 of 36 synthetic-corruption conditions, and on CheXpert it attains the best ACC, AUC, and F1 under both smoothing and U-zero. Controlled substitutions separate the roles of its components: prototype refinement supplies most of the resistance to label corruption, whereas the VQC yields the largest cosine margin, the smallest within-class covariance trace, and the highest refinement accuracy among the projectors tested. These empirical results agree with the recovery analysis, which relates correct prototype matching to trusted-anchor count, within-class covariance, and cosine margin; however, they support a projector-specific geometric advantage within the evaluated pipeline, not a claim of quantum speedup or general superiority over classical feature maps. Future work will compare trusted-data purification methods under matched anchor budgets and evaluate adaptive anchor acquisition, cross-institutional transfer, finite-shot estimation, and device noise.

AI USE STATEMENT

We used OpenAI's ChatGPT to assist with language editing, manuscript organization, and LaTeX formatting. The authors verified all scientific claims, analyses, results, citations, and final text.

REFERENCES

Naim Ajlouni, Adem Özyavaş, Mustafa Takaoğlu, Faruk Takaoğlu, and Firas Ajlouni. Medical image diagnosis based on adaptive hybrid quantum cnn. *BMC Medical Imaging*, 23(1):126, 2023.

Gamze Algan and Ilkay Ulusoy. Deep learning from small amount of medical data with noisy labels: A meta-learning approach. *arXiv preprint arXiv:2010.06939*, 2020.

Vladimir Aliev, Pavel Ostyakov, Roman Suvorov, Gleb Sterkin, Elizaveta Logacheva, Oleg Khomenko, and Sergey Nikolenko. Label denoising with large ensembles of heterogeneous neural networks. In *Proc. of the 2nd Workshop on YouTube-8M Large-Scale Video Understanding*, 2018.

Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 312–321. PMLR, 2019.

Denis Bokhan, Alena S. Mastiukova, Aleksey S. Boev, Dmitrii N. Trubnikov, and Aleksey K. Fedorov. Multiclass classification using quantum convolutional neural networks with hybrid quantum–classical learning. *Frontiers in Physics*, 10:1069985, 2022. doi: 10.3389/fphy.2022.1069985.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001.

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

Yuanhong Chen, Fengbei Liu, Hu Wang, Chong Wang, Yuyuan Liu, Yu Tian, and Gustavo Carneiro. BoMD: Bag of multi-label descriptors for noisy chest x-ray classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21284–21295, 2023.

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

Bidisha Dhara, Monika Agrawal, and Sumantra Dutta Roy. Multi-class classification using quantum transfer learning. *Quantum Information Processing*, 23(2):34, 2024. doi: 10.1007/s11128-023-04237-1.

Alex Díaz and Damian Steele. Analysis of classifiers robust to noisy labels. *arXiv preprint arXiv:2106.00274*, 2021.

Erik Englesson and Hossein Azizpour. Robust classification via regression for learning with noisy labels. In *ICLR 2024-The Twelfth International Conference on Learning Representations, Messe Wien Exhibition and Congress Center, Vienna, Austria, May 7-11t, 2024*, 2024.

Chuanwen Feng, Yilong Ren, and Xike Xie. OT-Filter: An optimal transport filter for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16164–16174, 2023.

Benoît Frénay and Michel Verleysen. Classification in the presence of label noise: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5):845–869, 2014. doi: 10.1109/TNNLS.2013.2292894.

Bo Han, Quanming Yao, and Xingrui Yu. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, pp. 8527–8537, 2018.

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

Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. Using trusted data to train deep networks on labels corrupted by severe noise. In *Advances in Neural Information Processing Systems*, volume 31, pp. 10477–10486, 2018.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 590–597, 2019. doi: 10.1609/aaai.v33i01.3301590.

Lie Ju, Xin Wang, Lin Wang, Dwarikanath Mahapatra, Xin Zhao, Quan Zhou, Tongliang Liu, and Zongyuan Ge. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE transactions on medical imaging*, 41(6):1533–1546, 2022.

Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis*, 65:101759, 2020.

Ria Khatoniar, Vishu Vyas, Videet Acharya, Anshul Saxena, Amit Saxena, Rahul Neiwal, and Kunal Korgaonkar. Quantum convolutional neural network for medical image classification: A hybrid model. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)*, pp. 1–5, 2024. doi: 10.1109/TQCEBT59414.2024.10545121.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18661–18673, 2020.

Jonas Landman, Natansh Mathur, Yun Yvonna Li, Martin Strahm, Skander Kazdaghli, Anupam Prakash, and Iordanis Kerenidis. Quantum methods for neural networks and application to medical image classification. *Quantum*, 6:881, 2022.

Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020.

Yifan Li, Hu Han, Shiguang Shan, and Xilin Chen. DISC: Learning from noisy labels via dynamic instance-specific selection and correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24070–24079, 2023.

Fei Liang, Rui Xu, and Wenjun Zhao. A hybrid quantum-classical neural network with deep residual learning. *Neural Networks*, 143:133–147, 2021.

Kevin J. Liang, Samrudhdhi B. Rangrej, Vladan Petrovic, and Tal Hassner. Few-shot learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9089–9098, 2022.

Junhua Liu, Kwan Hui Lim, Kristin L Wood, Wei Huang, Chu Guo, and He-Liang Huang. Hybrid quantum-classical convolutional neural networks. *Science China Physics, Mechanics & Astronomy*, 64(9):290311, 2021.

Moona Mazher, Abdul Qayyum, MKA Ahamed Khan, Steven Niederer, Mastaneh Mokayef, and CS Hassan. Hybrid classical and quantum deep learning models for medical image classification. *Journal of Healthcare Engineering*, 2024:1–12, 2024.

Curtis G. Northcutt, Lu Jiang, and Isaac L. Chuang. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70:1373–1411, 2021. doi: 10.1613/jair.1.12125.

Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1944–1952, 2017.

Maqsudur Rahman and Jun Zhuang. NQNN: Noise-aware Quantum Neural Networks for Medical Image Classification . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15972. Springer Nature Switzerland, September 2025.

Maqsudur Rahman and Jun Zhuang. QAPNet: A quantum-attentive patchwise network for robust medical image classification under noisy inputs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25065–25072, 2026.

Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4334–4343. PMLR, 2018.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019. doi: 10.1103/PhysRevLett.122.040504.

Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020. doi: 10.1103/PhysRevA.101.032308.

Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. *Advances in neural information processing systems*, 32, 2019.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4077–4087, 2017.

Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8135–8153, 2023. doi: 10.1109/TNNLS.2022.3152527.

Daiki Tanaka, Daiki Ikami, Toshihiko Yamasaki, and Kiyoharu Aizawa. Joint optimization framework for learning with noisy labels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5552–5560, 2018. doi: 10.1109/CVPR.2018.00582.

Yuanpeng Tu, Boshen Zhang, Yuxi Li, Liang Liu, Jian Li, Yabiao Wang, Chengjie Wang, and Cai Rong Zhao. Learning from noisy labels with decoupled meta label purifier. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19934–19943, 2023. doi: 10.1109/CVPR52729.2023.01909.

Cheng Xue, Lequan Yu, Pengfei Chen, Qi Dou, and Pheng-Ann Heng. Robust medical image classification from noisy labeled data with global and local representation guided co-training. *IEEE transactions on medical imaging*, 41(6):1371–1382, 2022.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.

Jiangchao Yao, Jiajie Wang, Ivor Tsang, Ya Zhang, Jun Sun, Chengqi Zhang, and Rui Zhang. Deep learning from noisy image labels with quality embedding. *arXiv preprint arXiv:1711.00583*, 2017. arXiv:1711.00583v1.

Kun Yi and Jianxin Wu. Probabilistic end-to-end noise correction for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7017–7025, 2019. doi: 10.1109/CVPR.2019.00718.

Table 7: Key notation shared by the main paper and the supplement.

Symbol	Meaning
$\boldsymbol{\phi}_e(\mathbf{X})$	Stabilized VQC measurement feature at epoch e.
$\mathcal{A}_c,\ m_c$	Trusted anchor set for class c and its cardinality.
$\mathbf{p}_{c,e}$	Algorithmic prototype for class c at epoch e.
$\boldsymbol{\mu}_{c,e}$	Population prototype under the fixed epoch-e feature map.
$\boldsymbol{\Sigma}_{c,e}$	Within-class covariance of $\boldsymbol{\phi}_e(\mathbf{X})$.
$\mathbf{t}_e(\mathbf{X})$	Multi-class prototype-refined soft target.
$\widehat{y}_e(\mathbf{X})$	Hard class induced by the refined target.
$\alpha_e(\mathbf{X})$	Maximum refined-target probability used as confidence.
$\mathcal{G}_{c,e}$	Confidence-gated set assigned to class c.
β, τ, λ_e	EMA rate, confidence threshold, and refined-target weight.
$\Gamma_e(\mathbf{X})$	Population cosine margin.
Δ_e	Maximum trusted-anchor prototype-estimation error.
$\mathcal{E}_{c,e}$	Error of the algorithmic prototype relative to the current population prototype.
$\zeta_{c,e}, \omega_{c,e}$	Gated-mean error and adjacent-epoch population drift.
$\eta, \rho_e, \eta_{\text{eff},e}$	Observed-label error, clean-class refined mass, and mixed-target incorrect mass.

Li Yi, Sheng Liu, Qi She, A. Ian McLeod, and Boyu Wang. On learning contrastive representations for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16682–16691, 2022.

Xiaoqing Ying, Hao Liu, and Rong Huang. Covid-19 chest x-ray image classification in the presence of noisy labels. *Displays*, 76:102370, 2023. doi: 10.1016/j.displa.2023.102370.

Mohammed Yousif, Belal Al-Khateeb, and Begonya Garcia-Zapirain. A new quantum circuits of quantum convolutional neural network for x-ray images classification. *IEEE Access*, 2024.

Yue Zhang, Yiyi Chen, Chaowei Fang, Qian Wang, Jiayi Wu, and Jingmin Xin. Learning from open-set noisy labels based on multi-prototype modeling. *Pattern Recognition*, 157:110902, 2025.

A Overview and Notation

This appendix provides the theoretical analysis (Apps. B–C), the training algorithm and circuit (App. D), the controlled-variant definitions (App. E), per-dataset controlled results (App. F), the trusted-anchor study (App. G), geometry and correction diagnostics (App. H), hyperparameter sweeps (App. I), and reproducibility details (App. J). Table 7 summarizes notation reused across sections; every other symbol is defined at first use.

B Assumptions

The clean-anchor result analyzes a fixed learned representation, while the EMA result separately accounts for representation drift during training.

Assumption 1 (Clean conditional anchors)**.** *For every class c, the images in $\mathcal{A}_c$ are conditionally independent draws from the clean class-conditional distribution, and their audited labels satisfy $y^\star = c$.*

Assumption 2 (Epoch-conditional feature map)**.** *For the trusted-anchor concentration and recovery analysis, the epoch-e feature map $\boldsymbol{\phi}_e$ is treated as fixed. Probability is taken over the trusted-anchor draw conditional on that learned feature map.*

Assumption 3 (Bounded and nondegenerate features)**.** *Every stabilized feature satisfies $\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \le 1$. For each image covered by the cosine-recovery statement, $\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 > 0$. Every population prototype satisfies $\|\boldsymbol{\mu}_{c,e}\|_2 \ge \kappa > 0$.*

Assumption 4 (Well-defined gated update)**.** *If $\mathcal{G}_{c,e}$ is empty, then $\mathbf{p}_{c,e+1} = \mathbf{p}_{c,e}$. Otherwise, $\overline{\boldsymbol{\phi}}_{c,e}$ is its empirical mean. The gated-mean error $\zeta_{c,e}$ and population drift $\omega_{c,e}$ are finite.*

C Complete Theoretical Analysis

C.1 Trusted-Anchor Concentration

For class c at epoch e, the trusted-anchor estimator is

$$\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}} = \frac{1}{m_c} \sum_{\mathbf{X} \in \mathcal{A}_c} \boldsymbol{\phi}_e(\mathbf{X}). \tag{16}$$

The superscript "anc" distinguishes this analytical estimator from the evolving algorithmic prototype $\mathbf{p}_{c,e}$.

Lemma 1 (Coordinate-wise Bernstein concentration)**.** *Let $\psi_\delta = \log(2QJ/\delta)$. Under Assumptions 1–3, with probability at least $1-\delta$, every class satisfies*

$$\left\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}} - \boldsymbol{\mu}_{c,e}\right\|_2 \le \sqrt{\frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_{c,e})\psi_\delta}{m_c}} + \frac{4\sqrt{Q}\,\psi_\delta}{3m_c}. \tag{17}$$

Proof. Fix class c and coordinate $k \in \{1, \dots, Q\}$. For anchor index a, let $Z_{a,k}$ denote the k-th coordinate of $\boldsymbol{\phi}_e(\mathbf{X}_a)$. Assumption S3 gives $Z_{a,k} \in [-1, 1]$, so the centered variable $Z_{a,k} - \mathbb{E}[Z_{a,k}]$ has absolute magnitude at most two and variance $[\boldsymbol{\Sigma}_{c,e}]_{kk}$. A two-sided scalar Bernstein inequality (Boucheron et al., 2013) gives

$$\Pr\left(\left|[\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}} - \boldsymbol{\mu}_{c,e}]_k\right| > \sqrt{\frac{2[\boldsymbol{\Sigma}_{c,e}]_{kk}\psi_\delta}{m_c}} + \frac{2\psi_\delta}{3m_c} \right) \le 2e^{-\psi_\delta}. \tag{18}$$

A union bound over Q coordinates and J classes makes the total failure probability at most δ. On the complementary event, Minkowski's inequality yields

$$\begin{aligned} \left\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}} - \boldsymbol{\mu}_{c,e}\right\|_2 &\le \left(\sum_{k=1}^{Q} \frac{2[\boldsymbol{\Sigma}_{c,e}]_{kk}\psi_\delta}{m_c} \right)^{1/2} \\ &\quad + \left(\sum_{k=1}^{Q} \frac{16\psi_\delta^2}{9m_c^2} \right)^{1/2} \qquad (19) \\ &= \sqrt{\frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_{c,e})\psi_\delta}{m_c}} \\ &\quad + \frac{4\sqrt{Q}\,\psi_\delta}{3m_c}. \qquad (20) \end{aligned}$$

which proves Eq. (17). □

C.2 Cosine Perturbation and Correct Refinement

Define the maximum anchor-estimation error

$$\Delta_e = \max_c \left\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}} - \boldsymbol{\mu}_{c,e}\right\|_2. \tag{21}$$

Lemma 2 (Prototype normalization)**.** *Suppose $\|\boldsymbol{\mu}_{c,e}\|_2 \ge \kappa$ and $\Delta_e < \kappa$. Then*

$$\left\| \frac{\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}}}{\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}}\|_2} - \frac{\boldsymbol{\mu}_{c,e}}{\|\boldsymbol{\mu}_{c,e}\|_2} \right\|_2 \le \frac{2\Delta_e}{\kappa - \Delta_e}. \tag{22}$$

The same quantity bounds the absolute perturbation of a cosine score computed with any unit feature direction.

Proof. The reverse triangle inequality gives $\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}}\|_2 \geq \kappa - \Delta_e > 0$. Add and subtract $\boldsymbol{\mu}_{c,e}/\|\widehat{\boldsymbol{\mu}}_{c,e}^{\mathrm{anc}}\|_2$, apply the triangle inequality, and use $|\|\widehat{\boldsymbol{\mu}}\|_2 - \|\boldsymbol{\mu}\|_2| \leq \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2$. The two resulting terms are each bounded by $\Delta_e/(\kappa - \Delta_e)$, proving Eq. (22). Taking an inner product with a unit direction cannot increase the norm difference. □

For an image $\mathbf{X}$ with latent clean class c, the population cosine margin is

$$\Gamma_e(\mathbf{X}) = \frac{\boldsymbol{\phi}_e(\mathbf{X})^\top \boldsymbol{\mu}_{c,e}}{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \|\boldsymbol{\mu}_{c,e}\|_2} - \max_{j \neq c} \frac{\boldsymbol{\phi}_e(\mathbf{X})^\top \boldsymbol{\mu}_{j,e}}{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2 \|\boldsymbol{\mu}_{j,e}\|_2}. \tag{23}$$

Theorem 1 (Clean-anchor prototype recovery)**.** *On the concentration event of Lemma 1, suppose* $\Delta_e < \kappa$ *and*

$$\Gamma_e(\mathbf{X}) > \frac{4\Delta_e}{\kappa - \Delta_e}. \tag{24}$$

Then empirical cosine prototype matching with the trusted-anchor means returns $y^\star$ *for* $\mathbf{X}$.

Proof. Lemma 2 bounds the perturbation of each class cosine score by $2\Delta_e/(\kappa - \Delta_e)$. The empirical gap between the latent clean class and the strongest competing class is therefore at least

$$\Gamma_e(\mathbf{X}) - \frac{4\Delta_e}{\kappa - \Delta_e}, \tag{25}$$

which is positive under Eq. (24). □

Corollary 1 (Sufficient trusted-anchor count)**.** *Let* $m_{\min} = \min_c m_c$ *and* $\nu_e = \max_c \mathrm{tr}(\boldsymbol{\Sigma}_{c,e})$. *Suppose every image targeted for guaranteed correction satisfies* $\Gamma_e(\mathbf{X}) > \Gamma_0 > 0$. *A sufficient condition for Theorem 1 is*

$$m_{\min} \geq \max\left\{ \frac{8\nu_e \psi_\delta (4+\Gamma_0)^2}{\kappa^2 \Gamma_0^2}, \frac{8\sqrt{Q}\,\psi_\delta (4+\Gamma_0)}{3\kappa\Gamma_0} \right\}. \tag{26}$$

For fixed Q, κ, *and* ν_e, *the dominant scaling is* $O(\log(J/\delta)/\Gamma_0^2)$.

Proof. Under Eq. (26), the square-root and linear terms in Eq. (17) are each at most $\kappa\Gamma_0/[2(4+\Gamma_0)]$. Hence $\Delta_e \leq \kappa\Gamma_0/(4+\Gamma_0) < \kappa$. Rearrangement gives $4\Delta_e/(\kappa - \Delta_e) \leq \Gamma_0 < \Gamma_e(\mathbf{X})$, so Theorem S1 applies. □

C.3 VQC-Induced Explicit-Feature Kernel

For a fixed epoch, define the stabilized unit direction

$$\mathbf{d}_e(\mathbf{X}) = \frac{\boldsymbol{\phi}_e(\mathbf{X})}{\max\{\|\boldsymbol{\phi}_e(\mathbf{X})\|_2, \xi_0\}} \tag{27}$$

and the explicit feature kernel

$$\mathcal{K}_e(\mathbf{X}, \mathbf{X}') = \mathbf{d}_e(\mathbf{X})^\top \mathbf{d}_e(\mathbf{X}'). \tag{28}$$

Proposition 4 (Positive semidefiniteness)**.** *For fixed epoch* e, $\mathcal{K}_e$ *is positive semidefinite.*

Proof. For any finite collection of images $\{\mathbf{X}_a\}$ and real coefficients $\{\gamma_a\}$,

$$\sum_{a,b} \gamma_a \gamma_b \mathcal{K}_e(\mathbf{X}_a, \mathbf{X}_b) = \left\| \sum_a \gamma_a \mathbf{d}_e(\mathbf{X}_a) \right\|_2^2 \geq 0. \tag{29}$$

□

The term "kernel" is used only in this explicit-feature-map sense. Equation (28) is not the quantum-state fidelity $|\langle\psi(\mathbf{X})|\psi(\mathbf{X}')\rangle|^2$.

C.4 Effective Incorrect Target Mass

Let $\eta = \Pr(y^{\text{obs}} \neq y^\star)$ be the observed-label error rate, and let $\rho_e = \mathbb{E}[t_{y^\star,e}(\mathbf{X})]$ be the expected refined-target mass placed on the latent clean class.

Proposition 5 (Mixed-target noise reduction; restates Proposition 2)**.** *The mixed target used by the multi-class objective has expected incorrect mass*

$$\eta_{\text{eff},e} = (1-\lambda_e)\eta + \lambda_e(1-\rho_e). \tag{30}$$

It is strictly smaller than η exactly when $\rho_e > 1-\eta$ and $\lambda_e > 0$.

Proof. Cross-entropy is linear in its target distribution. The expected mass assigned by the mixed target to the latent clean class is

$$(1-\lambda_e)(1-\eta) + \lambda_e\rho_e. \tag{31}$$

Subtracting this expression from one gives Eq. (30). The strict-improvement condition follows by direct rearrangement. □

For hard refined assignments, $\rho_e = \Pr(\widehat{y}_e(\mathbf{X}) = y^\star)$. For a CheXpert observation indexed by o, define its policy-dependent observed-target error rate η_o and expected refined positive-target correctness $\rho_{o,e}$. The same derivation gives

$$\eta_{\text{eff},o,e} = (1-\lambda_e)\eta_o + \lambda_e(1-\rho_{o,e}). \tag{32}$$

C.5 EMA Stability with Representation Drift

Define

$$\mathcal{E}_{c,e} = \|\mathbf{p}_{c,e} - \boldsymbol{\mu}_{c,e}\|_2, \quad \zeta_{c,e} = \|\overline{\boldsymbol{\phi}}_{c,e} - \boldsymbol{\mu}_{c,e}\|_2, \tag{33}$$

and

$$\omega_{c,e} = \|\boldsymbol{\mu}_{c,e+1} - \boldsymbol{\mu}_{c,e}\|_2. \tag{34}$$

Proposition 6 (One-step and unrolled EMA bounds; extends Proposition 3)**.** *Under Assumption 4,*

$$\mathcal{E}_{c,e+1} \leq (1-\beta)\mathcal{E}_{c,e} + \beta\zeta_{c,e} + \omega_{c,e}. \tag{35}$$

Unrolling the recursion gives

$$\mathcal{E}_{c,e} \leq (1-\beta)^e\mathcal{E}_{c,0} + \sum_{\ell=0}^{e-1}(1-\beta)^{e-1-\ell}\left(\beta\zeta_{c,\ell} + \omega_{c,\ell}\right). \tag{36}$$

Proof. Substitute the EMA update $\mathbf{p}_{c,e+1} = (1-\beta)\mathbf{p}_{c,e} + \beta\overline{\boldsymbol{\phi}}_{c,e}$, add and subtract $\boldsymbol{\mu}_{c,e}$, and apply the triangle inequality to obtain Eq. (35). Repeated substitution proves Eq. (36). □

If the feature map is frozen and $\zeta_{c,e} \leq \overline{\zeta}_c$, then

$$\limsup_{e\to\infty} \mathcal{E}_{c,e} \leq \overline{\zeta}_c. \tag{37}$$

If $\omega_{c,e} \leq \overline{\omega}_c$ also holds, the steady-state bound becomes $\overline{\zeta}_c + \overline{\omega}_c/\beta$. Thus, τ primarily affects the contamination of the admitted set through $\zeta_{c,e}$, whereas β trades preservation of the trusted initialization against tracking of a changing feature map.

D Complete Algorithmic Specification

Algorithm 1 expands the compact description in the main paper. It explicitly separates gradient-based model updates from detached refined targets and epoch-level prototype updates.

CheXpert specialization. For each CheXpert observation, audited certain-negative and certain-positive examples initialize two prototypes. A two-class softmax over the corresponding cosine scores produces one component of $\mathbf{q}_e^{\text{ref}}(\mathbf{X})$. Uncertain entries are excluded from anchor construction. The confidence rule, empty-gate behavior, numerical stabilization, and detachment operations are applied independently to every observation.

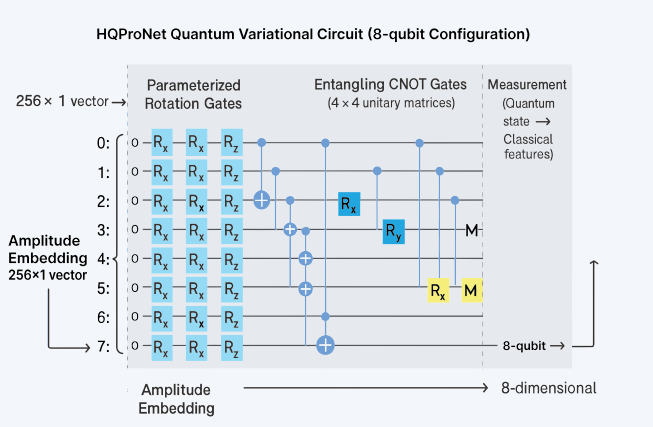


Figure 4: **Eight-qubit, five-layer VQC used by HQProNet.** The normalized 256-dimensional amplitude vector is encoded into eight qubits. Each trainable layer applies single-qubit rotations followed by ring-CNOT entanglement. Per-qubit Pauli-Z expectations form the eight-dimensional measured feature used for classification and prototype matching.

Algorithm 1 HQProNet Training with Trusted-Anchor Refinement

Require: Noisy training data; trusted anchor sets $\{\mathcal{A}_c\}_{c=1}^{J}$; confidence threshold τ; EMA rate β; temperature T; refined-target schedule λ_e.

Initialize g_{enc}, g_{qual}, $\mathbf{W}_{\text{amp}}$, the VQC, and the classifier.
Encode the trusted anchors and initialize $\mathbf{p}_{c,0} = m_c^{-1} \sum_{\mathbf{X}\in\mathcal{A}_c} \boldsymbol{\phi}_0(\mathbf{X})$ for each class c.
for each epoch e **do**
  for each training minibatch **do**
    Compute $\mathbf{h}(\mathbf{X})$, $\mathbf{v}(\mathbf{X})$, $\mathbf{z}(\mathbf{X})$, $\mathbf{r}_e(\mathbf{X})$, and $\boldsymbol{\phi}_e(\mathbf{X})$.
    Compute the prototype scores, the multi-class target $\mathbf{t}_e(\mathbf{X})$ or CheXpert target $\mathbf{q}_e^{\text{ref}}(\mathbf{X})$, the confidence $\alpha_e(\mathbf{X})$, and the classifier output.
    Evaluate $\mathcal{L}_e^{\text{mc}}$ or $\mathcal{L}_e^{\text{ml}}$ and update only the trainable model parameters; detach refined targets and prototypes from gradient computation.
  end for
  Form each $\mathcal{G}_{c,e}$ from the full-epoch hard assignments and confidence values.
  If $\mathcal{G}_{c,e}$ is nonempty, compute $\overline{\boldsymbol{\phi}}_{c,e}$ and apply the EMA update; otherwise retain $\mathbf{p}_{c,e}$.
end for
return Trained encoder, quality module, amplitude projection, VQC, classifier, and prototype bank.

Degenerate and empty cases. If $\mathbf{v}(\mathbf{X}) = \mathbf{0}$, the implementation maps the sample to a fixed computational-basis state before the VQC. If a stabilized feature or prototype has a very small norm, ξ_0 is used in the denominator. If no sample is admitted for a class in an epoch, its prototype remains unchanged.

E Controlled Variants and Protocol

The component ablation (Table 3) additionally uses Naïve, which disables both the VQC and the prototype mechanism and is trained only with the observed-label objective; its Linear+Proto row is the variant of the same name below. Table 8 fixes the operational definition of every control.

All projector-based variants receive the same 256-dimensional amplitude-projection input and produce an eight-dimensional normalized feature. They use identical trusted anchors, corruption masks, classifier, optimizer, refinement schedule, early stopping, and tuning budget. Classical-only is retained as a lower-bound control and therefore classifies directly from the shared 80-dimensional feature.

For MedMNIST, corruption is applied only to training labels, and every method receives the same

Table 8: Exact controlled-variant definitions. Projector parameter counts exclude the shared encoder and classifier.

Variant	Projector and retained components	Output	Params	Prototype refinement
Classical-only	Classify directly from the shared 80-dimensional hybrid feature; remove the eight-dimensional projector and prototype module.	80	0	No
Linear+Proto	Apply a trainable affine $256 \to 8$ map followed by Euclidean normalization; retain trusted prototypes, confidence gating, EMA, and the mixed objective.	8	2,056	Yes
MLP+Proto	Apply a trainable $256 \to 32 \to 8$ GELU MLP followed by Euclidean normalization; retain the complete prototype mechanism.	8	8,488	Yes
Budget-MLP+Proto	Apply a fixed orthogonal $256 \to 8$ compression followed by a trainable $8 \to 6 \to 8$ GELU MLP; retain the complete prototype mechanism.	8	110	Yes
VQC-only	Use the eight-qubit, five-layer VQC and classifier; disable refined targets and EMA updates by setting $\lambda_e = 0$.	8	120	No
HQProNet	Use the same VQC with trusted prototypes, confidence gating, EMA updates, and mixed supervision.	8	120	Yes

saved corruption realization for a given seed. For CheXpert, the 10% patient-level subset is formed before splitting, patient disjointness is preserved, and uncertain entries are handled by smoothing or U-zero rather than interpreted as a known flip process. The trusted-budget experiment yields anchor-count ranges of 1–11, 2–21, 6–53, and 12–106 per class at budgets of 1%, 2%, 5%, and 10%, respectively.

F Extended Controlled Results

The main paper summarizes the controlled projector study through averages across datasets and corruption settings. This section reports the corresponding per-dataset means and standard deviations so that the consistency of the projector comparison can be examined directly. Every controlled variant follows the operational definition in Table 8 and uses the same encoder, classifier, data partitions, corruption realization, trusted anchor draw, prototype-update rule, optimizer, stopping criterion, and tuning budget. Consequently, the reported differences isolate the replacement of the projector and, for VQC-only, the removal of prototype-guided refinement.

The analysis has three parts. First, we examine controlled projectors under symmetric corruption, where corrupted labels are distributed across incorrect classes. Second, we consider pairwise corruption, where repeated class-to-class transitions can systematically bias a prototype. Third, we evaluate smoothing and U-zero on CheXpert, where uncertainty handling is policy dependent rather than generated by a known label-flip process. Tables 9–13 report mean macro metrics with seed-level standard deviations.

F.1 Symmetric-Noise Projector Results

Symmetric corruption replaces a selected training label with an incorrect class without favoring one particular transition. The 30% setting evaluates substantial corruption, while the 50% setting tests whether the projector and refinement mechanism remain reliable when half of the training labels are altered. Tables 9 and 10 divide the six datasets into two groups only for readability: Table 9 contains BloodMNIST, PathMNIST, and OrganAMNIST, whereas Table 10 contains OrganCMNIST, DermaMNIST, and OCTMNIST.

Across all dataset–corruption combinations in Tables 9 and 10, HQProNet obtains the highest mean macro-F1. Averaged across the six datasets, HQProNet reaches 80.2% at 30% corruption and 78.9%

Table 9: Macro-F1 (%) for BloodMNIST, PathMNIST, and OrganAMNIST under 30% and 50% symmetric training-label corruption.

Variant	Blood 30	Blood 50	Path 30	Path 50	OrganA 30	OrganA 50
Linear+Proto	81.1 ± 0.6	80.2 ± 0.7	83.7 ± 0.5	81.2 ± 0.6	73.8 ± 0.8	74.6 ± 0.7
MLP+Proto	83.0 ± 0.5	80.8 ± 0.6	83.7 ± 0.4	82.4 ± 0.5	78.3 ± 0.6	77.3 ± 0.7
Budget-MLP+Proto	82.3 ± 0.6	80.1 ± 0.7	83.0 ± 0.5	81.7 ± 0.6	77.6 ± 0.7	76.6 ± 0.8
VQC-only	74.9 ± 0.8	73.0 ± 0.9	79.2 ± 0.7	75.6 ± 0.8	71.2 ± 0.9	65.8 ± 1.0
HQProNet	$\mathbf{85.1 \pm 0.4}$	$\mathbf{83.2 \pm 0.5}$	$\mathbf{85.8 \pm 0.4}$	$\mathbf{84.8 \pm 0.5}$	$\mathbf{80.4 \pm 0.5}$	$\mathbf{79.7 \pm 0.6}$

Table 10: Macro-F1 (%) for OrganCMNIST, DermaMNIST, and OCTMNIST under 30% and 50% symmetric training-label corruption.

Variant	OrganC 30	OrganC 50	Derma 30	Derma 50	OCT 30	OCT 50
Linear+Proto	74.6 ± 0.7	72.8 ± 0.8	72.2 ± 0.7	73.0 ± 0.8	72.9 ± 0.8	71.1 ± 0.9
MLP+Proto	76.7 ± 0.6	75.3 ± 0.7	74.3 ± 0.6	72.9 ± 0.7	72.4 ± 0.7	70.7 ± 0.8
Budget-MLP+Proto	76.0 ± 0.7	74.6 ± 0.8	73.7 ± 0.7	72.2 ± 0.8	71.7 ± 0.8	70.0 ± 0.9
VQC-only	65.9 ± 0.9	63.9 ± 1.0	62.1 ± 1.0	58.8 ± 1.1	66.5 ± 0.9	62.1 ± 1.0
HQProNet	$\mathbf{78.8 \pm 0.5}$	$\mathbf{77.7 \pm 0.6}$	$\mathbf{76.4 \pm 0.5}$	$\mathbf{75.2 \pm 0.6}$	$\mathbf{74.4 \pm 0.6}$	$\mathbf{73.0 \pm 0.7}$

at 50% corruption. The strongest classical projector, MLP+Proto, obtains 78.1% and 76.6%, respectively. The corresponding HQProNet improvements are therefore 2.1 and 2.4 macro-F1 points.

The per-dataset pattern is also consistent. At 30% corruption, the HQProNet improvement over MLP+Proto lies between 2.0 and 2.1 points across the six datasets. At 50% corruption, the improvement remains between 2.3 and 2.4 points. Thus, the aggregate advantage in the main paper is not produced by one unusually favorable dataset. The reported standard deviations for HQProNet remain between 0.4 and 0.7, indicating that the direction of the comparison is stable over the three repeated runs.

The classical controls clarify the source of the improvement. MLP+Proto generally improves over Linear+Proto, showing that nonlinear projection can benefit prototype matching, although this improvement is not uniform on DermaMNIST and OCTMNIST at 50% corruption. Budget-MLP+Proto remains competitive despite its smaller trainable parameter count, but is consistently below HQProNet. VQC-only averages 70.0% and 66.5% at 30% and 50% corruption, respectively. The substantial difference between VQC-only and HQProNet shows that the measured VQC representation alone is not sufficient: the trusted initialization, refined supervision, confidence gate, and EMA updates are essential components of the complete method.

F.2 Pairwise-Noise Projector Results

Pairwise corruption maps each selected class to a prespecified competing class. Unlike symmetric corruption, the resulting errors repeatedly favor the same transition and can therefore move a class prototype toward a particular incorrect class. Tables 11 and 12 repeat the controlled projector comparison under this more structured setting. The dataset grouping is identical to that used for the symmetric results, allowing direct comparison between corruption mechanisms.

HQProNet remains strongest in all pairwise-corruption conditions. Its six-dataset averages are 76.8% at 30% corruption and 75.6% at 50% corruption. MLP+Proto obtains 74.7% and 73.2%, giving HQProNet improvements of 2.2 and 2.4 macro-F1 points. Linear+Proto reaches 73.1% and 71.7%, while Budget-MLP+Proto reaches 74.0% and 72.5%. The ordering therefore remains broadly consistent across corruption rates and projector capacities.

The pairwise results are lower than the corresponding symmetric results for every controlled variant, which is consistent with the additional difficulty created by repeated class-dependent transitions. Nevertheless, HQProNet retains a positive advantage on every dataset. Its improvement over MLP+Proto ranges from 2.1 to 2.5 points across the individual pairwise conditions, rather than being concentrated in one dataset or one corruption rate.

Table 11: Macro-F1 (%) for BloodMNIST, PathMNIST, and OrganAMNIST under 30% and 50% pairwise training-label corruption.

Variant	Blood 30	Blood 50	Path 30	Path 50	OrganA 30	OrganA 50
Linear+Proto	77.6 ± 0.7	76.2 ± 0.8	80.0 ± 0.6	78.1 ± 0.7	70.7 ± 0.9	70.1 ± 0.9
MLP+Proto	79.4 ± 0.6	77.2 ± 0.7	80.0 ± 0.5	78.7 ± 0.6	74.8 ± 0.7	74.0 ± 0.8
Budget-MLP+Proto	78.7 ± 0.7	76.5 ± 0.8	79.3 ± 0.6	78.0 ± 0.7	74.2 ± 0.8	73.3 ± 0.9
VQC-only	71.2 ± 0.9	69.3 ± 1.0	75.4 ± 0.8	72.1 ± 0.9	67.8 ± 1.0	62.2 ± 1.1
HQProNet	$\mathbf{81.6 \pm 0.5}$	$\mathbf{79.7 \pm 0.6}$	$\mathbf{82.2 \pm 0.4}$	$\mathbf{81.2 \pm 0.5}$	$\mathbf{77.0 \pm 0.6}$	$\mathbf{76.4 \pm 0.7}$

Table 12: Macro-F1 (%) for OrganCMNIST, DermaMNIST, and OCTMNIST under 30% and 50% pairwise training-label corruption.

Variant	OrganC 30	OrganC 50	Derma 30	Derma 50	OCT 30	OCT 50
Linear+Proto	71.6 ± 0.8	69.8 ± 0.9	69.1 ± 0.8	68.4 ± 0.9	69.5 ± 0.9	67.7 ± 1.0
MLP+Proto	73.4 ± 0.7	72.0 ± 0.8	71.1 ± 0.7	69.7 ± 0.8	69.2 ± 0.8	67.6 ± 0.9
Budget-MLP+Proto	72.7 ± 0.8	71.3 ± 0.9	70.4 ± 0.8	69.0 ± 0.9	68.5 ± 0.9	66.9 ± 1.0
VQC-only	62.6 ± 1.0	60.4 ± 1.1	58.5 ± 1.1	55.1 ± 1.2	62.9 ± 1.0	58.8 ± 1.1
HQProNet	$\mathbf{75.5 \pm 0.6}$	$\mathbf{74.4 \pm 0.7}$	$\mathbf{73.2 \pm 0.6}$	$\mathbf{72.1 \pm 0.7}$	$\mathbf{71.3 \pm 0.7}$	$\mathbf{70.0 \pm 0.8}$

VQC-only averages 66.4% at 30% pairwise corruption and 63.0% at 50%, substantially below the complete model. This result again separates feature projection from supervision repair: the VQC can define a useful compact geometry, but without trusted prototypes and refined targets it cannot prevent systematic corrupted labels from continuing to influence the classifier. The complete result is therefore best interpreted as the interaction between the measured VQC feature map and the prototype-refinement mechanism.

F.3 CheXpert Projector Comparison

CheXpert differs from the controlled MedMNIST experiments because its uncertain entries do not provide a known latent label-flip process. We therefore compare the projectors under two uncertainty-handling policies rather than assigning a corruption percentage. Smoothing retains uncertain entries as soft supervision, whereas U-zero maps them to the negative class. Table 13 reports macro-averaged accuracy, AUC, and F1 across the five selected clinical observations. These three metrics distinguish overall decision agreement, ranking performance, and class-balanced predictive quality.

HQProNet obtains the highest mean value for all six reported CheXpert metrics. Under smoothing, it reaches 90.3% accuracy, 91.5% AUC, and 77.2% macro-F1. Relative to MLP+Proto, these values correspond to gains of 2.2, 2.1, and 2.2 points, respectively. Under U-zero, HQProNet reaches 85.3% accuracy, 81.0% AUC, and 70.6% macro-F1, improving over MLP+Proto by 2.5, 1.7, and 2.6 points.

Macro-F1 is lower under U-zero for every projector, indicating that mapping uncertainty directly to the negative class produces a more demanding class-balanced setting. The decrease is particularly large for Linear+Proto, whose macro-F1 changes from 74.8% under smoothing to 63.9% under U-zero. HQProNet also decreases, but remains the strongest variant under both policies. Because latent clean labels for uncertain CheXpert entries are unavailable, these results evaluate predictive robustness and do not support direct repair-precision or repair-recall claims.

Figure 5 places the four principal F1 conditions from the main-paper projector comparison on a common axis: the six-dataset averages at 50% symmetric and pairwise corruption and CheXpert macro-F1 under smoothing and U-zero. The figure is intended to show whether the projector ordering remains consistent across synthetic corruption and policy-dependent clinical uncertainty; the complete metric values remain in Tables 9–13.

The profile shows three complementary effects. First, Classical-only is substantially below every projector-based variant, confirming that the shared representation without the controlled eight-dimensional projector and refinement pipeline is an insufficient lower bound. Second, VQC-only

Table 13: CheXpert macro metrics (%) under smoothing (Sm) and U-zero (U0).

Variant	Sm ACC	Sm AUC	Sm F1	U0 ACC	U0 AUC	U0 F1
Linear+Proto	87.8 ± 0.4	89.1 ± 0.4	74.8 ± 0.5	82.0 ± 0.5	79.2 ± 0.6	63.9 ± 0.7
MLP+Proto	88.1 ± 0.4	89.4 ± 0.3	75.0 ± 0.4	82.8 ± 0.5	79.3 ± 0.5	68.0 ± 0.5
Budget-MLP+Proto	87.5 ± 0.5	88.8 ± 0.4	74.3 ± 0.5	82.2 ± 0.5	78.9 ± 0.6	67.2 ± 0.6
VQC-only	85.5 ± 0.5	87.0 ± 0.5	72.9 ± 0.6	80.5 ± 0.6	77.4 ± 0.7	65.6 ± 0.7
HQProNet	$\mathbf{90.3 \pm 0.3}$	$\mathbf{91.5 \pm 0.3}$	$\mathbf{77.2 \pm 0.4}$	$\mathbf{85.3 \pm 0.4}$	$\mathbf{81.0 \pm 0.5}$	$\mathbf{70.6 \pm 0.5}$

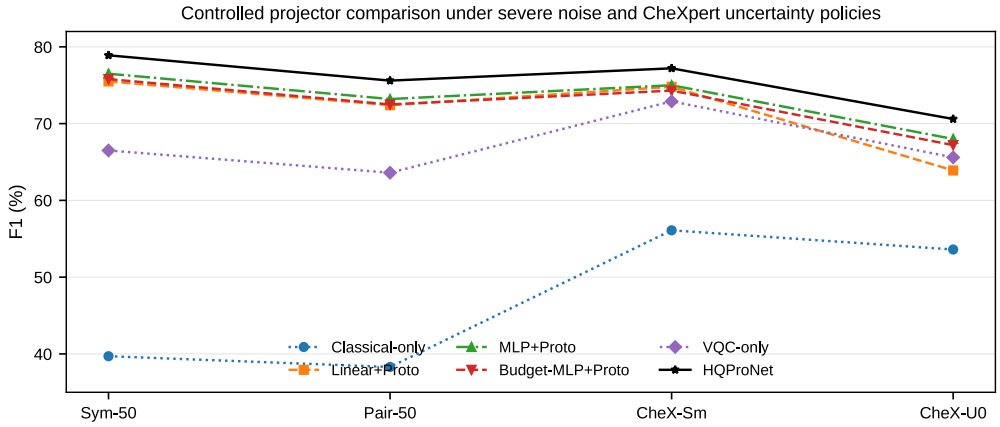


Figure 5: Predictive profile of the controlled projectors. HQProNet is strongest across severe symmetric corruption, severe pairwise corruption, and both CheXpert uncertainty policies.

improves over Classical-only but remains below all prototype-based controls, again showing that projection and supervision repair play distinct roles. Third, HQProNet is highest in all four conditions. Its margin over MLP+Proto is moderate but consistent, supporting the bounded claim that the VQC measurement map provides a complementary benefit when the prototype mechanism and training protocol are held fixed.

G Trusted-Anchor Sensitivity

The trusted-anchor analysis varies only the fraction of the validation-class examples used to initialize the prototype bank. Budgets of 1%, 2%, 5%, and 10% are evaluated with the architecture, corruption realizations, optimization settings, confidence rule, and EMA procedure held fixed. Because the datasets and classes have different cardinalities, a common percentage produces different absolute anchor counts. The realized range is 1–11 examples per class at 1%, 2–21 at 2%, 6–53 at 5%, and 12–106 at 10%. This distinction is important because the theoretical concentration result depends on the absolute class-wise count m_c, not on the percentage alone.

Table 14 first summarizes the six-dataset averages and correction behavior. It jointly reports predictive performance, corrupted-label repair, and clean-label retention so that an apparent F1 gain cannot be interpreted without considering how aggressively the method changes labels.

Increasing the trusted budget from 1% to 5% improves macro-F1 by 2.5 points at 30% symmetric corruption, 3.0 points at 50% symmetric corruption, 2.2 points at 30% pairwise corruption, and 2.3 points at 50% pairwise corruption. Over the same interval, repair accuracy increases from 79.6% to 85.8%, while clean-label retention increases from 93.2% to 95.1%. The improvement therefore reflects both more successful correction of corrupted labels and slightly better preservation of labels that were already correct.

The returns diminish beyond the default setting. Increasing the budget from 5% to 10% adds only 0.4 points at 30% symmetric corruption and 0.3 points in each of the other three corruption conditions. Repair increases by 0.7 points and retention by 0.2 points. These changes are considerably smaller than those obtained between 1% and 5%, despite doubling the trusted annotation budget.

Table 14: Average macro-F1 and label-repair behavior versus trusted annotation budget. Repair is the percentage of corrupted labels restored to the latent clean class; retention is the percentage of initially clean labels not changed incorrectly.

Budget	Count range/class	Sym-30	Sym-50	Pair-30	Pair-50	Repair	Retention
1%	1–11	76.8 ± 0.8	75.1 ± 1.0	73.8 ± 0.9	72.5 ± 1.1	79.6 ± 1.2	93.2 ± 0.6
2%	2–21	78.2 ± 0.7	76.6 ± 0.8	74.8 ± 0.8	73.6 ± 0.9	82.3 ± 1.0	94.1 ± 0.5
5%	6–53	79.3 ± 0.5	78.1 ± 0.6	76.0 ± 0.6	74.8 ± 0.7	85.8 ± 0.8	95.1 ± 0.4
10%	12–106	79.7 ± 0.5	78.4 ± 0.6	76.3 ± 0.6	75.1 ± 0.7	86.5 ± 0.7	95.3 ± 0.4

Table 15: Per-dataset HQProNet macro-F1 at 50% corruption. In each cell, the top line is symmetric and the bottom line is pairwise.

Budget	Blood	Path	OrganA	OrganC	Derma	OCT
	79.2 ± 1.0	81.0 ± 0.9	75.7 ± 1.1	73.5 ± 1.2	71.4 ± 1.1	70.1 ± 1.2
1%	75.9 ± 1.1	77.5 ± 1.0	72.7 ± 1.2	71.1 ± 1.2	68.6 ± 1.2	69.0 ± 1.3
	80.9 ± 0.8	82.3 ± 0.7	77.2 ± 0.9	75.0 ± 0.9	73.1 ± 0.9	71.3 ± 1.0
2%	77.2 ± 0.9	78.8 ± 0.8	74.1 ± 1.0	72.2 ± 1.0	70.1 ± 1.0	69.0 ± 1.1
	82.4 ± 0.5	84.0 ± 0.5	78.9 ± 0.6	76.9 ± 0.6	74.4 ± 0.6	72.3 ± 0.7
5%	78.9 ± 0.6	80.4 ± 0.5	75.6 ± 0.7	73.7 ± 0.7	71.4 ± 0.7	69.3 ± 0.8
	82.7 ± 0.5	84.2 ± 0.4	79.2 ± 0.6	77.2 ± 0.6	74.7 ± 0.6	72.5 ± 0.7
10%	79.2 ± 0.6	80.7 ± 0.5	75.9 ± 0.7	74.0 ± 0.7	71.7 ± 0.7	69.5 ± 0.8

The averages in Table 14 could conceal dataset-specific behavior. Table 15 therefore reports the 50% symmetric and pairwise results for each dataset separately. The upper entry in each cell is the symmetric result and the lower entry is the pairwise result. The per-dataset values show that the plateau is not caused by averaging improvements and decreases across datasets. Between 1% and 5%, symmetric-noise macro-F1 increases by 2.2–3.4 points across all six datasets. Under pairwise corruption, five datasets increase by 2.6–3.0 points, while OCTMNIST increases by only 0.3 points, suggesting that its pairwise setting is comparatively insensitive to additional anchors over this range. From 5% to 10%, however, every dataset changes by only 0.2 or 0.3 points under each corruption type.

Absolute performance also remains dataset dependent. BloodMNIST and PathMNIST achieve the largest values across all budgets, whereas DermaMNIST and OCTMNIST remain more challenging. The trusted-budget conclusion should therefore be understood as a consistent saturation trend rather than as evidence that all datasets reach the same performance level. Figure 6 visualizes the entries of Table 15 as symmetric- and pairwise-corruption heatmaps. The common intensity scale makes both the monotonic budget trend and the persistent differences in dataset difficulty directly visible.

The heatmaps show the largest visual change between the 1% and 5% rows, followed by comparatively small changes between 5% and 10%. We therefore select 5% as an empirical annotation–performance operating point. This choice is not theoretically unique: the sufficient condition in Sec. C.2 depends jointly on absolute anchor count, within-class covariance, prototype norm, failure probability, and cosine margin. A different dataset or representation may consequently require a different trusted fraction.

H Theory-Linked Geometry and Correction Diagnostics

The prototype-recovery result in Secs. C.2–C.4 identifies two representation properties that directly affect the sufficient condition: a larger cosine margin and a smaller within-class covariance. The experiments therefore measure both quantities together with the behavior of the resulting label-refinement process.

Refinement accuracy is the fraction of all training examples whose hard refined assignment equals the latent clean label. Repair precision is the fraction of changed observed labels that are changed to the latent clean class. Repair recall is the fraction of corrupted labels restored to the latent clean class. Retention is the fraction of initially clean labels that are not changed incorrectly. The covariance

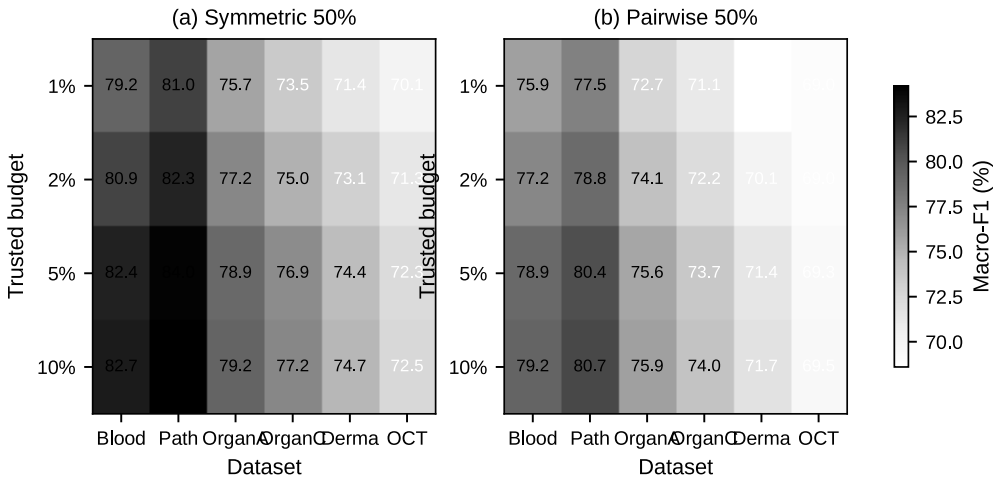


Figure 6: Per-dataset trusted-budget sensitivity at 50% symmetric and pairwise corruption. Improvements are strongest between 1% and 5%, after which most datasets approach a plateau.

Table 16: Diagnostics at 50% symmetric corruption. Higher margin, refinement accuracy, repair precision, repair recall, and retention are better; lower covariance trace is better. Percentage-valued diagnostics are reported in %.

Variant	Margin	Cov. trace	Refine acc.	Repair precision	Repair recall	Retention
Linear+Proto	0.180 ± 0.011	0.244 ± 0.014	83.7 ± 0.8	80.3 ± 1.0	76.5 ± 1.1	94.1 ± 0.5
MLP+Proto	0.214 ± 0.010	0.217 ± 0.012	85.2 ± 0.7	82.7 ± 0.9	79.4 ± 1.0	94.8 ± 0.4
Budget-MLP+Proto	0.201 ± 0.012	0.226 ± 0.013	84.5 ± 0.8	81.8 ± 0.9	78.3 ± 1.0	94.6 ± 0.5
VQC-only	0.194 ± 0.013	0.230 ± 0.015	–	–	–	–
HQProNet	$\mathbf{0.246 \pm 0.009}$	$\mathbf{0.185 \pm 0.011}$	$\mathbf{88.0 \pm 0.6}$	$\mathbf{86.3 \pm 0.7}$	$\mathbf{83.8 \pm 0.8}$	$\mathbf{95.8 \pm 0.4}$

trace is weighted by class frequency, and the cosine margin is averaged over the corresponding analysis set.

Classical-only is omitted from these tables because it does not use the common eight-dimensional projector and prototype-refinement space. VQC-only has well-defined margin and covariance values, but its refinement, repair, and retention entries are marked with dashes because refined targets and prototype updates are disabled. These dashes therefore denote quantities that are not applicable, rather than missing experimental runs.

Table 16 reports the diagnostics under 50% symmetric corruption. Table 17 reports the corresponding results under 50% pairwise corruption.

Under symmetric corruption, HQProNet has a mean cosine margin of 0.246 and a covariance trace of 0.185. Relative to MLP+Proto, the strongest classical projector, the margin increases by 0.032 while the covariance trace decreases by 0.032. The corresponding refinement accuracy increases from 85.2% to 88.0%, repair precision from 82.7% to 86.3%, repair recall from 79.4% to 83.8%, and retention from 94.8% to 95.8%.

The simultaneous increase in repair recall and retention is important. It indicates that the higher correction rate is not obtained simply by modifying more labels indiscriminately: HQProNet restores a larger fraction of corrupted labels while also preserving a larger fraction of initially correct labels.

Pairwise corruption makes the geometry less favorable for every projector. For example, HQProNet's margin decreases from 0.246 to 0.221, while its covariance trace increases from 0.185 to 0.201. The same direction is visible for MLP+Proto, whose margin changes from 0.214 to 0.189 and whose covariance trace changes from 0.217 to 0.233. This shift is consistent with the additional difficulty caused by systematic class-to-class corruption.

Table 17: Diagnostics at 50% pairwise corruption. Percentage-valued diagnostics are reported in %.

Variant	Margin	Cov. trace	Refine acc.	Repair precision	Repair recall	Retention
Linear+Proto	0.154 ± 0.013	0.259 ± 0.016	81.0 ± 0.9	77.4 ± 1.1	72.8 ± 1.2	93.5 ± 0.6
MLP+Proto	0.189 ± 0.011	0.233 ± 0.014	82.9 ± 0.8	80.1 ± 1.0	76.0 ± 1.1	94.1 ± 0.5
Budget-MLP+Proto	0.177 ± 0.012	0.241 ± 0.015	82.2 ± 0.9	79.3 ± 1.0	75.1 ± 1.1	93.9 ± 0.5
VQC-only	0.169 ± 0.014	0.247 ± 0.017	–	–	–	–
HQProNet	$\mathbf{0.221 \pm 0.010}$	$\mathbf{0.201 \pm 0.012}$	$\mathbf{85.6 \pm 0.7}$	$\mathbf{84.0 \pm 0.8}$	$\mathbf{80.9 \pm 0.9}$	$\mathbf{95.0 \pm 0.4}$

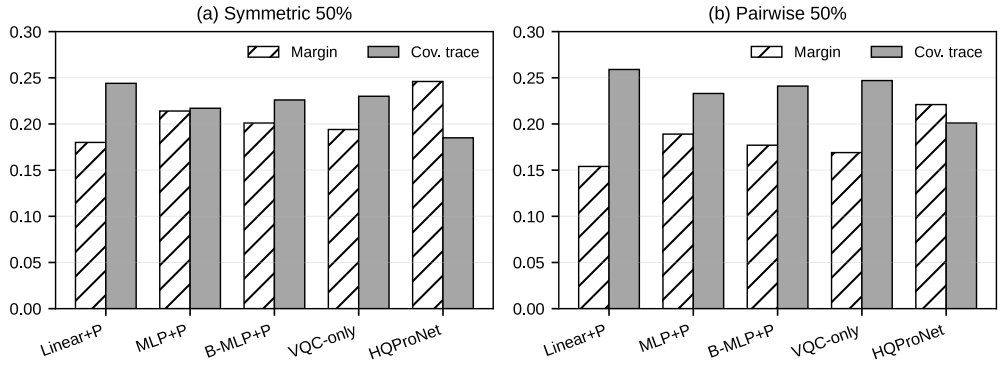


Figure 7: Margin and covariance-trace comparison under 50% symmetric and pairwise corruption. HQProNet combines the largest mean separation with the lowest within-class dispersion among the controlled projectors.

Despite this shift, HQProNet retains the strongest values. Compared with MLP+Proto, it increases the pairwise margin by 0.032, lowers the covariance trace by 0.032, improves refinement accuracy by 2.7 points, repair precision by 3.9 points, repair recall by 4.9 points, and retention by 0.9 points. The similar direction of the symmetric and pairwise results supports the consistency of the geometry and correction observations.

Figure 7 isolates margin and covariance trace because these are the two measured quantities that enter most directly into the prototype-recovery analysis. The bars should be interpreted comparatively: larger margin is favorable, whereas smaller covariance trace is favorable.

Across both panels, HQProNet occupies the favorable high-margin, low-dispersion position. VQC-only also exhibits a more favorable geometry than some lower controls, but its absence of refinement prevents that geometry from producing label-repair measurements. Conversely, the classical prototype variants perform correction but do not reach the same combination of separation and concentration as the complete model.

I Hyperparameter Stability

The main paper presents a PathMNIST case study, whereas this section evaluates the same hyperparameter families using the primary MedMNIST metric averaged across all six datasets. The analysis varies one component at a time around the default configuration $\lambda_{\max} = 0.5$, $T = 1.0$, $\beta = 0.2$, and $\tau = 0.7$. Figure 8 first provides a visual overview of the one-dimensional sweeps and the two-dimensional EMA/confidence grid. Tables 18 and 19 then report the exact means and standard deviations.

The curves indicate a local plateau rather than an isolated optimum. Values adjacent to the selected defaults produce similar macro-F1, while larger deviations reduce both predictive performance and refinement accuracy. The heatmap shows that the EMA rate and confidence threshold also interact: an update rate that is acceptable with a selective gate can become less stable when the gate admits lower confidence assignments. Table 18 quantifies the refined-target-weight and temperature sweeps. The refined-target ceiling controls how strongly the prototype target contributes relative to the observed label, whereas temperature controls the sharpness of the prototype-derived distribu-

Table 18: Sensitivity of the refined-target ceiling $\lambda_{\max}$ and temperature T. Macro-F1 and refinement accuracy are reported in %.

Sweep	Value	Sym-50 F1	Pair-50 F1	Refine acc.
$\lambda_{\max}$	0.3	77.3 ± 0.6	74.1 ± 0.7	86.0 ± 0.7
$\lambda_{\max}$	0.4	77.8 ± 0.5	74.5 ± 0.7	87.1 ± 0.6
$\lambda_{\max}$	**0.5**	$\mathbf{78.1 \pm 0.5}$	$\mathbf{74.8 \pm 0.6}$	$\mathbf{88.0 \pm 0.6}$
$\lambda_{\max}$	0.6	77.9 ± 0.6	74.6 ± 0.7	87.9 ± 0.6
$\lambda_{\max}$	0.7	77.4 ± 0.7	74.2 ± 0.8	87.2 ± 0.7
T	0.5	77.4 ± 0.7	74.2 ± 0.8	87.4 ± 0.7
T	**1.0**	$\mathbf{78.1 \pm 0.5}$	$\mathbf{74.8 \pm 0.6}$	$\mathbf{88.0 \pm 0.6}$
T	1.5	77.8 ± 0.6	74.5 ± 0.7	87.6 ± 0.6
T	2.0	77.2 ± 0.7	74.0 ± 0.8	86.9 ± 0.7
T	3.0	76.4 ± 0.8	73.2 ± 0.9	85.8 ± 0.8

Table 19: Macro-F1 (%) at 50% symmetric corruption for the EMA-rate and confidence-threshold grid.

β/τ	0.60	0.70	0.80
0.10	77.1 ± 0.7	77.6 ± 0.6	77.3 ± 0.7
0.20	77.8 ± 0.6	$\mathbf{78.1 \pm 0.5}$	77.7 ± 0.6
0.40	75.6 ± 0.8	77.2 ± 0.7	76.9 ± 0.8

tion. For $\lambda_{\max}$, the selected value of 0.5 obtains 78.1% symmetric-noise F1, 74.8% pairwise-noise F1, and 88.0% refinement accuracy. The neighboring values 0.4 and 0.6 remain close, differing by at most 0.3 macro-F1 points, which supports local stability. Reducing the ceiling to 0.3 underweights the refined target and decreases symmetric/pairwise F1 to 77.3/74.1. Increasing it to 0.7 also reduces performance to 77.4/74.2, indicating that the observed label should not be discarded too aggressively.

Temperature exhibits a similar pattern. The default $T = 1.0$ is strongest, while $T = 0.5$ and $T = 1.5$ remain within 0.7 and 0.3 symmetric-noise F1 points, respectively. At $T = 3.0$, the refined distribution becomes substantially more diffuse and macro-F1 falls to 76.4% under symmetric and 73.2% under pairwise corruption. Refinement accuracy decreases in the same direction, from 88.0% at $T = 1.0$ to 85.8% at $T = 3.0$.

Table 19 examines the interaction between the EMA update rate β and confidence threshold τ. The threshold controls which assignments are admitted to the prototype update, while the EMA rate determines how strongly the admitted mean changes the existing prototype.

The default pair $(\beta, \tau) = (0.2, 0.7)$ obtains the highest mean macro-F1, 78.1%. A threshold of 0.7 is also the strongest value within each tested EMA row, suggesting that the middle confidence level provides a useful balance between coverage and selectivity. The lowest result, 75.6%, occurs for the most aggressive tested update rate, $\beta = 0.4$, combined with the most permissive threshold, $\tau = 0.6$. This pattern is consistent with rapid prototype movement being vulnerable when lower-confidence samples are admitted.

The surrounding configurations nevertheless remain competitive: $(0.1, 0.7)$ gives 77.6% and $(0.2, 0.8)$ gives 77.7%. The selected configuration should therefore be viewed as a stable operating point within the evaluated grid, not as evidence that performance depends on one precisely tuned value. Taken together, the sweeps support the design choices used in the main experiments. Moderate refined-target weighting, temperature near one, a selective confidence threshold, and a non-aggressive EMA update produce the most reliable combination of macro-F1 and refinement accuracy. These results are empirical and do not imply that the same defaults are universally optimal outside the evaluated datasets and corruption protocols.

Figure 8 summarizes both one-dimensional sweeps and the two-dimensional (β, τ) grid.

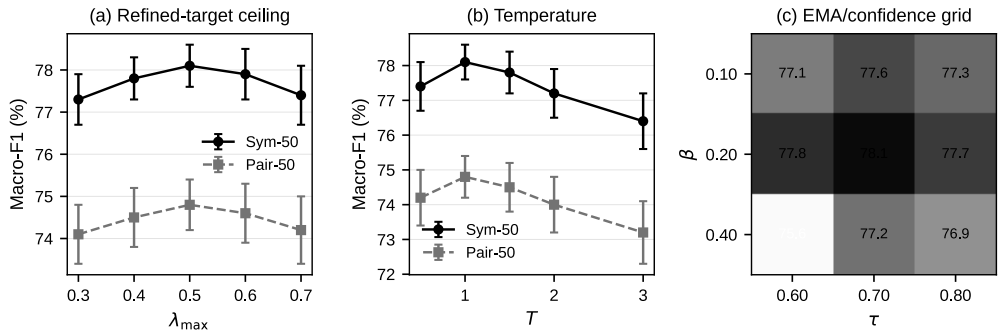


Figure 8: Macro-F1 sensitivity using the paper's primary MedMNIST metric. The defaults $\lambda_{\max} = 0.5$, $T = 1.0$, $\beta = 0.2$, and $\tau = 0.7$ lie in a stable region.

The aggregate results support a broad local plateau rather than a single fragile setting. Smaller $\lambda_{\max}$ underweights the refined target, very large T makes it too diffuse, and the fastest EMA update is less stable when the confidence threshold is permissive.

J Reproducibility and Claim Boundaries

Statistical reporting. Tables 9–19 report mean $\pm$ standard deviation over three training seeds. In particular, HQProNet is strongest in every controlled MedMNIST projector condition and in every reported CheXpert metric. The standard deviations describe seed-level variation; no claim of statistical significance is made beyond the reported repeated-run evidence.

Controlled projector attribution. The projector study holds fixed the shared encoder, classifier, data split, corruption mask, trusted-anchor draw, prototype initialization and update rules, optimizer, refined target schedule, early-stopping criterion, and tuning budget. Linear+Proto, MLP+Proto, Budget-MLP+Proto, and HQProNet also use the same eight-dimensional output size. These controls permit a specific comparison among affine, nonlinear classical, approximately parameter-matched classical, and VQC measurement projectors.

Trusted supervision boundary. HQProNet assumes access to a small audited anchor set. These anchors are disjoint from the model-selection subset, are supplied identically to all prototype-based controls, and do not contribute gradients. The method is therefore robust learning with limited trusted supervision rather than fully unsupervised label-noise correction. The trusted-budget results characterize the measured annotation–performance trade-off for the evaluated splits.

Synthetic corruption and CheXpert. For MedMNIST, the latent clean labels remain available after synthetic corruption, allowing refinement accuracy, repair precision, repair recall, and retention to be computed directly. CheXpert is different: smoothing and U-zero are alternative policies for uncertain report-derived entries, and a known latent flip process is unavailable. Consequently, the CheXpert experiments support claims about predictive robustness under uncertainty handling.

Quantum scope. The VQC experiments are conducted through classical simulation. The eight-qubit, five-layer circuit contains 120 trainable rotations, but parameter count alone does not establish reduced wall-clock cost, quantum speedup, or hardware advantage. The defensible conclusion is narrower: within the controlled simulated setting, the VQC measurement map provides consistent predictive, geometric, and correction gains when compared with the evaluated classical projectors. The positive-semidefinite function of App. C.3 is an explicit kernel over measured features.

Experiment provenance. Reproducibility requires retaining the exact patient- or sample-level data partitions, corruption masks, trusted-anchor draws, model-selection splits, random seeds, hyperparameter configurations, and software versions. Per-seed predictions, refined targets, hard assignments, confidence values, prototype states, and correction indicators provide the inputs needed to reconstruct the reported means and standard deviations.