Quantum-Enhanced Federated Learning via QAOA Feature and Gate Selection with Skew-Symmetric Lipschitz Dynamics

Mahmudul Hasan *, Sharif Ullah †, Sachin Shetty *
*Old Dominion University, Norfolk, Virginia, USA
†University of Central Arkansas, Conway, AR, USA
Email: mhasa005@odu.edu, mullah@uca.edu, sshetty@odu.edu

***Abstract*—Federated learning (FL) trains models across many clients without sharing raw data, but large feature sets and non-IID client data can slow training and increase communication. We introduce a Quantum–Lipschitz Federated Network (QLFN) that pairs fixed-size selection with stable training. LightGBM first ranks features, and a Quantum Approximate Optimization Algorithm (QAOA) based selector then chooses a fixed-size subset. A second QAOA step prunes hidden gates using gradient scores. To keep learning stable under heterogeneity, QLFN adds a skew-symmetric Lipschitz flow that limits gradient growth. On the UCI Cardiotocography dataset with a Dirichlet non-IID split across 10 clients, the QLFN and LightGBM ensemble achieves 92.3% test accuracy and F1 score is 0.85, lifting recall for the Pathologic class to 0.91 with similar communication cost. The selectors settle early and refresh only occasionally, yielding targeted pruning and regrowth rather than full mask changes. Overall, combining QAOA based selection with Lipschitz control improves robustness, interpretability, and efficiency for tabular federated learning.**

***Index Terms*—Federated learning, QAOA, quantum feature map, Lipschitz flow, skew-symmetric dynamics, LightGBM, distributed optimization.**

I. Introduction

Federated learning (FL) decentralizes model training across multiple clients without sharing raw data, which is a critical property for privacy-sensitive domains. Practically the non-IID, noisy, and redundant client data can slow convergence, increase communication cost, and hinder interpretability. Recent tabular benchmarks indicate that neural networks are sensitive to irrelevant inputs and that feature selection (FS) should be evaluated by downstream neural performance rather than proxy metrics [1]. This motivates FS methods that are noise-aware, communication-efficient, interpretable, and aligned with model behavior. High dimensional heterogeneity further degrades Federated learning by increasing communication cost and aggregation inefficiency: [2] show that noisy local features exacerbate non-IID aggregation and overhead, while [3] report that redundancy raises communication cost and lowers global accuracy.

High-dimensional, noisy, and heterogeneous data further reduce FL performance by inflating communication cost and slowing convergence. Liu et al. [2] show that, in non-IID settings, noisy local features drive inefficient aggregation and higher communication overhead; correspondingly, reducing feature dimensionality can improve privacy and scalability. Similarly, Banerjee et al. [3] report that feature redundancy increases communication cost and decreases global accuracy for heterogeneous clients at scale.

In this paper we introduce the Quantum-Lipschitz Federated Network (QLFN), which incorporates FS directly into the federated pipeline instead of a preprocessing step. Firstly, feature-gain scores are generated by a compact tree model (LightGBM), and then converted into a quadratic selection objective and investigated using a quantum-inspired QAOA selector. Here, a cardinality constraint k is enforced to remove noisy features, reduce communication cost, and increase stability. Where k is the exact number of input features, a smaller k filters noise and decreases communication cost, while a larger k retains more signal. The selected indices are written on the interface of the model; for this reason, the network encodes the raw columns. That means the feature selection (FS) has a direct link with the model and is not an external filter. A skew-symmetric Lipschitz flow block is created from the skew sections of normalized learned matrices. This block take inputs after they have been mapped through a QAOA feature map. At the same time as. maintaining expressive dynamics, this flow stabilizes representations. QLFN computes gradient-based gate importances on a validation slice. It periodically reselects a binary gate mask with the same QAOA objective. This allows QLFN to adapt capacity under non-IID heterogeneity. As a result, targeted activation and pruning are possible without requiring architectural modifications. The federated scaffold provides training and testing metrics for the quantum model, the LightGBM baseline, and their ensemble. It supports standard optimizers, such as FedAvg and FedProx. In addition, it supports Top-K update sparsification and reports train/test metrics for the quantum model, the LightGBM baseline, and their ensemble.

In short, QLFN links LightGBM-based relevance estimation to QAOA-guided, cardinality-constrained feature selection. It embeds this selection directly into the model's input path. The method pairs this step with a skew-symmetric Lipschitz flow and periodic gate reselection. This approach balances interpretability, robustness, and communication efficiency under non-IID FL.

II. Related Work

Modern studies states that evaluation should be performance-driven with deep downstream learners, and they document regimes where neural models are sensitive to noise compared to strong GBDT baselines [1]. These findings motivate FS stages that protect deep models from extraneous variables—precisely the role FS plays in our QLFN front end.

Hybrid FS pipelines typically filter first and then run a wrapper. A persistent gap is choosing the number k of retained features. Treating k as a decision variable, recent work optimizes a task-aware FS score via Bayesian Optimization or Golden Section Search, yielding larger reductions and lower runtime at parity accuracy, and further accelerating downstream wrappers [4]. We adopt the same principle—optimize discrete cardinality—by embedding it directly in a QAOA Ising objective, giving a physics-motivated way to enforce k in a single stage.

Generative views of FS encode subsets into permutation-invariant embeddings and then search that space with a reinforcement-learning policy to balance accuracy and subset length, mitigating permutation bias and non-convexity pitfalls [5]. While our method does not learn a full subset-embedding space, it echoes that philosophy: we combine a learned classical prior with a principled discrete optimizer (QAOA) and refresh masks during training, which acts like a guided search over sparse structures.

In HFL, one line of work targets secure/lightweight filters: SeiFS reduces expensive comparisons via custom encodings and oblivious access and employs approximate top-k selection, yielding order-of-magnitude speedups [2]. A complementary HFL direction optimizes the accuracy–cost trade-off: Fed-MOFS couples MI with clustering and a multiobjective global ranking, reporting $\approx 50\%$ feature-space reduction with $O(d^2)$ complexity and faster training than prior HFL FS [3]. Beyond filter/hybrid designs, embedded FFS integrates selection during training: DSFFS employs dynamic sparse training that prunes/regrows input and hidden connections online, improving the accuracy–communication–compute balance under non-IID splits [6]. In VFL, FedSDG-FS introduces stochastic dual gates with partially homomorphic encryption and a Gini-based importance initializer, enabling secure embedded selection with low communication and strong downstream performance [7].

Our approach differs in two respects: (1) we enforce an explicit cardinality constraint for both inputs and latent gates via QAOA, unifying cutoff selection and subset choice; and (2) we stabilize optimization with a skew-symmetric Lipschitz flow. This complements FedAvg/FedProx and coexists with privacy mechanisms (e.g., DP) and compression, aiming to reduce redundancy and improve minority-class recall under realistic non-IID splits.

III. Dataset

We evaluated QLFN using biomedical datasets from the UCI Machine Learning Repository to assess robustness under imbalanced and heterogeneous data conditions.

A. Cardiotocography (CTG) Dataset

The Cardiotocography (CTG) dataset [8] consists of 2,126 fetal monitoring records collected from cardiotocogram (CTG) sensors. Each record includes 21 numerical features computed from fetal heart rate (FHR) and uterine contraction (UC) signals. In addition, Samples are labeled into three clinical categories:

- Normal: 1655 samples,
- Suspect: 295 samples,
- Pathologic:176 samples.

The dataset is highly imbalanced, with approximately 78% of the samples belonging to the Normal class, making rare-class recognition (Pathologic cases) particularly challenging. We normalize all features to the range $[0, 1]$ and split the dataset 80/20 for training and testing. For federated experiments, data are partitioned among 10 clients using a Dirichlet distribution ($\alpha = 0.5$) to simulate non-IID distributions.

IV. Methodology

In Fig 1 describes a schematic of the QAOA-based quantum circuit used for feature or gate selection. The circuit begins with an initial state ψ and applies sequential parameterized rotations (RX, RY, RZ) on each qubit, followed by controlled entangling operations to capture correlations between qubits. After several such alternating layers, measurements yield bitstrings that represent optimized feature or gate selections for the federated model. In addition, Fig. 2 refers to the overall process.

A. QAOA for Selection and Encoding

a) From adiabatic evolution to a Trotterized circuit.: We start from a time–dependent evolution generated by the sum of a cost Hamiltonian H_C and a mixer $H_M = \sum_{j=1}^{n} X_j$. A first order Lie–Trotter product over p slices yields the (discrete) QAOA ansatz

$$\psi(\boldsymbol{\gamma},\boldsymbol{\beta}) = \left[\prod_{\ell=1}^{p} e^{-i\beta_\ell H_M}\, e^{-i\gamma_\ell H_C}\right] +^{\otimes n}, \quad \boldsymbol{\gamma},\boldsymbol{\beta} \in \mathbb{R}^p, \tag{1}$$

and the parameters are obtained by the variational principle

$$(\boldsymbol{\gamma}^\star,\boldsymbol{\beta}^\star) \in \arg\min_{\boldsymbol{\gamma},\boldsymbol{\beta}} \underbrace{\psi(\boldsymbol{\gamma},\boldsymbol{\beta})|H_C|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})}_{\mathcal{L}(\boldsymbol{\gamma},\boldsymbol{\beta})}. \tag{2}$$

Equations (1)–(2) define the quantum state we optimize.

b) Cardinality-penalized Ising for discrete selection: We want a k-hot binary mask over d items with importance weights $w \in \mathbb{R}^d$. The constrained selection is relaxed into a penalty:

$$E_{\text{bit}}(b) = -\,w^\top b + \lambda\Big(\sum_{i=1}^{d} b_i - k\Big)^2, \qquad b \in \{0,1\}^d. \tag{3}$$

Introduce spins $s_i \in \{-1,+1\}$ via $b_i = \frac{1-s_i}{2}$ and expand:

$$E_{\text{spin}}(s) = -\,w^\top\left(\frac{1-s}{2}\right) + \lambda\left(\frac{1}{2}\sum_{i=1}^{d}(1-s_i) - k\right)^2$$
$$= \sum_{i=1}^{d}\alpha_i\, s_i + \sum_{i<j}\beta\, s_i s_j + \text{const},$$

Where,

- $\alpha_i = -\,w_i - \frac{\lambda d}{2} + \lambda k$
- $\beta = \frac{\lambda}{2}$
- $w \in \mathbb{R}^d$: weights; $s_i \in \{-1,+1\}$: spins; k: target cardinality; d: number of items

Identifying Pauli operators Z_i with spin variables gives the Ising cost Hamiltonian

$$H_C = \sum_{i=1}^{d}\alpha_i Z_i + \sum_{i<j}\beta\, Z_i Z_j, \tag{4}$$

$$\alpha_i = -w_i - \frac{\lambda d}{2} + \lambda k, \tag{5}$$

$$\beta = \frac{\lambda}{2}. \tag{6}$$

Eqns (3)–(4) show how a k-cardinality problem becomes an Ising Hamiltonian that QAOA can minimize.

c) Selector readout and cardinality guarantee.: After optimizing (2), we measure computational-basis outcomes as *bits* $b \in \{0,1\}^d$ and score them with the same bit-energy used in the objective,

$$E_{\text{readout}}(b) \;=\; -\,w^\top b \;+\; \lambda\left(\sum_{i=1}^{d} b_i - k\right)^2. \tag{7}$$

We then choose the best-scoring sample $b^\star \in \arg\min_b E_{\text{readout}}(b)$. (Equivalently, if one prefers to map measurements to spins $s_i = 1 - 2b_i \in \{-1,+1\}$, the same score can be written as $E_{\text{readout}}(s) = \frac{1}{2}\sum_i w_i s_i + \frac{\lambda}{4}\big(\sum_i s_i - (d-2k)\big)^2 + \text{const}$, which yields the same argmin.) To enforce exact cardinality when sampling misses k, we apply the fallback

$$\hat{z} = \begin{cases} b^\star, & \sum_i b_i^\star = k, \\ \text{TopK}(w,k), & \text{otherwise}, \end{cases} \tag{8}$$

which guarantees a valid k-hot mask at every selection step.

d) QAOA as Feature-Map: With n selected features, a learnable symmetric interaction $W = W^\top \in \mathbb{R}^{n\times n}$, a ring mask M, and layer angles $\{(\gamma_\ell,\beta_\ell)\}_{\ell=1}^{p}$, one layer applies

$$U_{\text{ent}}(\gamma_\ell, W) = \prod_{j<k}\exp\big(-\,i\,2\gamma_\ell (M\odot W)_{jk} Z_j Z_k\big), \tag{9}$$

$$U_{\text{data}}(\gamma_\ell, x) = \prod_{j=1}^{n}\exp\big(-\,i\,2\gamma_\ell x_j Z_j\big), \tag{10}$$

$$U_{\text{mix}}(\beta_\ell) = \prod_{j=1}^{n}\exp\big(-\,i\,2\beta_\ell X_j\big), \tag{11}$$

so the encoder state is

$$\phi(x) = \Big[\prod_{\ell=1}^{p} U_{\text{mix}}(\beta_\ell)\, U_{\text{data}}(\gamma_\ell, x)\, U_{\text{ent}}(\gamma_\ell, W)\Big] +^{\otimes n}. \tag{12}$$

Expectations produce features

$$\phi(x) = \begin{bmatrix} \langle Z_1\rangle, \ldots, \langle Z_n\rangle \\ \langle Z_1 Z_2\rangle, \ldots, \langle Z_{n-1} Z_n\rangle \end{bmatrix} \in \mathbb{R}^{n+\binom{n}{2}}. \quad (13)$$

$$\tilde{\phi}(x) = \mathrm{BN}\big(\phi(x)\big). \quad (14)$$

As illustrated in Fig. 1, the QAOA selector minimizes the Ising objective in (4) and the encoder layer applies (9)–(11) to produce the embedding (13).

B. LightGBM Baseline and Importance-Driven Seeding

We incorporate a classical gradient-boosted decision tree model (LightGBM) in two complementary ways: (i) to *seed* the QAOA feature selector with data-driven importances, and (ii) as a strong classical baseline that we ensemble with the quantum path. This makes the pipeline pragmatic: QAOA handles discrete cardinality selection using the Ising objective, while LightGBM supplies a stable prior over features and a point of comparison.

a) Gain-based importances.: Let $\mathcal{F} = \{1, \ldots, d\}$ index raw input features and I_j denote the cumulative split gain assigned by LightGBM to feature j (summed over all trees and nodes). We normalize the gains to form a nonnegative weight vector $\hat{w} \in \mathbb{R}^d$:

$$\hat{w}_j \;=\; \frac{I_j}{\max_{k\in\mathcal{F}} I_k + \varepsilon}, \qquad j \in \mathcal{F}, \quad (15)$$

with a small ε for numerical stability. These weights are used solely as a prior and do not contribute gradients to the quantum path.

b) Seeding QAOA feature selection.: The k-hot selection problem is relaxed into the cardinality-penalized Ising energy $E_{\text{bit}}(b)$ in (3), with the Ising coefficients in (4). To bias selection toward features with larger classical signal, we set $w\hat{w}$ in (3) and solve

$$b^\star \in \arg\min_{b\in\{0,1\}^d} \left\{ -\,\hat{w}^\top b + \lambda\Big(\sum_{i=1}^{d} b_i - k\Big)^2 \right\}, \quad (16)$$

using the QAOA ansatz in (1) and the variational objective in (2). The sampling-and-fallback rule in (**??**) guarantees an exact k-hot mask. In our default implementation we take K_{feat}=6; this can be refreshed periodically but is typically computed once at initialization.

c) Mapping to the encoder width.: The QAOA encoder consumes exactly n inputs (qubits). From the binary mask $m_{\text{feat}} \in \{0,1\}^d$ an ordered index buffer that maps raw columns to the encoder. If $K_{\text{feat}} > n$, we truncate to n entries; if $K_{\text{feat}} < n$, we pad by repeating the first entry so that the encoder always receives n values. The quantum embedding $\tilde{\phi}(x)$ used downstream is given by (13)–(14).

d) Evaluation and ensembling.: At evaluation time we report three models: (i) the quantum path, (ii) the LightGBM baseline, and (iii) a simple uniform probability ensemble of the two:

$$\hat{p}_{\text{ens}}(y \mid x) = \alpha\, \hat{p}_{\text{lgb}}(y \mid x) + (1-\alpha)\, \hat{p}_{\text{qnn}}(y \mid x) \quad (17)$$

Here, $\alpha \in [0,1]$ is the ensemble weight . with α=0.5 by default. We report accuracies on both train and test splits, and a confusion matrix and classification report for the ensemble.

e) Federated learning variant.: In the federated setting we compute $\hat{w}$ on a server-side held-out slice, run QAOA once at initialization to produce m_{feat} and the corresponding $\hat{z}_{\text{feat}}$, and broadcast these masks alongside gate masks to clients. LightGBM itself is *not* federated; it is used only to initialize the feature prior and to provide a consistent classical reference for comparison/ensembling across rounds. Periodic reseeding of features is optional and, when enabled, follows the same procedure with refreshed $\hat{w}$.

f) Gate bottleneck (classical).: From the QAOA embedding $\tilde{\phi}(x) \in \mathbb{R}^{n+\binom{n}{2}}$ we form a width-d_z latent

$$z_0 \;=\; W_{\text{lin}}\, \tilde{\phi}(x) + b_{\text{lin}} \;\in\; \mathbb{R}^{d_z}, \quad (18)$$

map it into a g-dimensional gate space

$$z_{\text{gate}} \;=\; W_{\text{z}\to\text{gate}}\, z_0 \;\in\; \mathbb{R}^{g}, \quad (19)$$

apply the binary gate mask $m \in \{0,1\}^g$ chosen by the QAOA selector,

$$\tilde{z}_{\text{gate}} \;=\; m \odot z_{\text{gate}}, \quad (20)$$

and project back to the flow width to obtain the modulation used by the Lipschitz flow:

$$z \;=\; W_{\text{gate}\to\text{z}}\, \tilde{z}_{\text{gate}} \;\in\; \mathbb{R}^{d_z}. \quad (21)$$

In the gate-importance computation (Eq. (27)), the sensitivities $\partial\ell/\partial z_{\text{gate}}$ are taken with respect to (19); these gradients backpropagate through the skew-symmetric flow (Eqs. (25)–(26)), yielding a well-conditioned importance signal for QAOA-based cardinality selection.

C. Skew-Symmetric Lipschitz Flow

[9] Let $h_t \in \mathbb{R}^d$ be the hidden state and $z \in \mathbb{R}^d$ the gate modulation. With trainable $B, C \in \mathbb{R}^{d\times d}$, form antisymmetric parts and their normalized versions:

$$B_{\text{skew}} = B - B^\top, \qquad \widehat{B}_{\text{skew}} = \frac{B_{\text{skew}}}{\|B_{\text{skew}}\| + \varepsilon},$$
$$C_{\text{skew}} = C - C^\top, \qquad \widehat{C}_{\text{skew}} = \frac{C_{\text{skew}}}{\|C_{\text{skew}}\| + \varepsilon}. \quad (22)$$

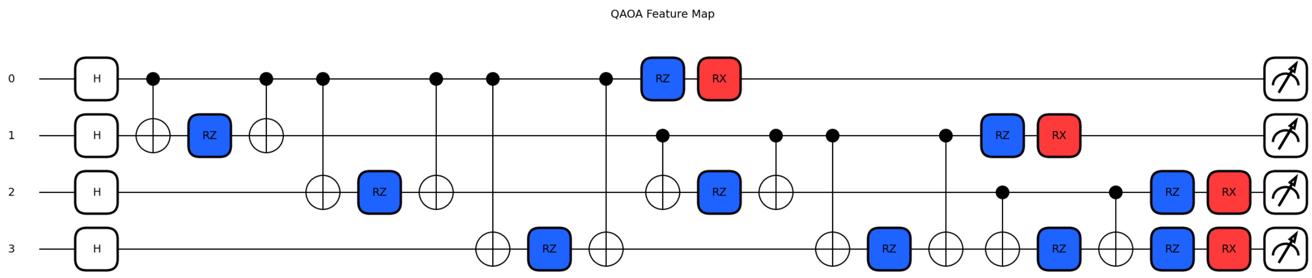

Fig. 1. Schematic of the QAOA selector/encoder stack used for feature/gate selection and embedding.

Blend normalized/raw skew parts and add damping:

$$A = \beta\,\widehat{B}_{\text{skew}} + (1-\beta)\,B_{\text{skew}} - \gamma I, \tag{23}$$

$$W = \beta\,\widehat{C}_{\text{skew}} + (1-\beta)\,C_{\text{skew}} - \gamma I. \tag{24}$$

A stable forward Euler step with batch norm and a mild nonlinearity is

$$r_t = \text{BN}\Big(h_t A + \text{LeakyReLU}(h_t W + z)\Big), \tag{25}$$

$$h_{t+1} = h_t + \Delta t \cdot r_t, \qquad t = 0, \ldots, T-1. \tag{26}$$

Equations. (23)–(26) define a contraction biased dynamics: antisymmetric cores act like rotations (energy–preserving), while $-\gamma I$ damps expansions, yielding a bounded Lipschitz constant for the post-quantum block.

a) Why this matters for selection.: Gate importances used by the selector are computed as layerwise sensitivity of the loss to the gate activations,

$$w_{\text{gate}} = \frac{1}{\big\| \sum_b |\partial\ell/\partial z_{\text{gate}}| \big\|_2 + \varepsilon} \sum_b \Big| \frac{\partial \ell}{\partial z_{\text{gate}}} \Big|. \tag{27}$$

Because gradients backpropagate through (25)–(26) and the Jacobians inherit the antisymmetry+damping structure, the signal in (27) is well-conditioned, which makes the subsequent QAOA selection more stable.

D. Federated Learning

a) Client sampling and masks.: At round t, the server samples S_t with $|S_t| = \lfloor CK \rfloor$. The current *feature* mask $\hat{z}_{\text{feat}}$ and *gate* mask $\hat{m}_{\text{gate}}$ are broadcast; clients restrict inputs and channels accordingly.

b) Local objective and encoder–flow path.: Each client i minimizes

$$\mathcal{L}_i^{\text{prox}}(\theta) = \frac{1}{|\mathcal{D}_i|} \sum_{(x,y)\in\mathcal{D}_i} \ell\Big(f_\theta(\tilde{\phi}(x)), y\Big) + \frac{\mu}{2}\,\|\theta - \theta_g^t\|_2^2, \tag{28}$$

where $\tilde{\phi}(x)$ is given by (13)–(14) and the flow is (25)–(26). Equation (28) ties the quantum encoder and the flow to the FL objective while controlling client drift.

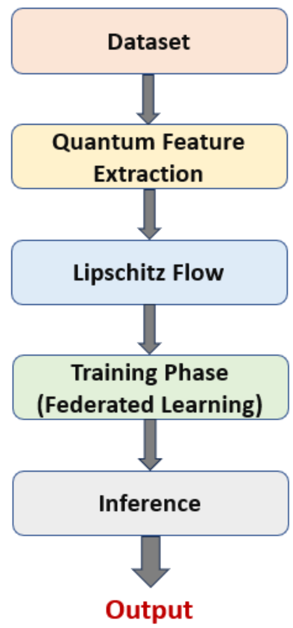

Fig. 2. **System workflow.**

c) Aggregation: Server parameters are updated by a size–weighted average:

$$\theta_g^{t+1} = \sum_{i\in S_t} \frac{n_i}{\sum_{j\in S_t} n_j}\,\theta_i^t. \tag{29}$$

d) Periodic selection: Every P_{gate} rounds P_{feat}), the server recomputes w_{gate} by (27) on a held-out slice and runs QAOA with the Ising form (4), then applies the cardinality rule to update masks.

V. Results

A. Federated Setup

We trained for 25 rounds with a client fraction of $\approx 30\%$ (3 clients per round). Gate masks were refreshed at rounds 1, 6, 11, 16, and 21:

- r1: [1 0 1 0 0 1 1 0]

TABLE I
OVERALL ACCURACY (TRAIN/TEST)

Model	Train Accuracy	Test Accuracy
QLFN	0.9924	0.9249

TABLE II
CLASS-WISE PRECISION/RECALL/F1 ON THE TEST SET (ENSEMBLE).

Class	Precision	Recall	F1
Normal	0.94	0.98	0.96
Suspect	0.84	0.63	0.72
Pathologic	0.84	0.91	0.88

- r6: [1 0 1 1 0 1 0 0]
- r11: [0 0 1 1 0 0 1 1]
- r16: [0 0 0 1 1 1 0 1]
- r21: [0 1 0 1 1 0 0 1]

B. Overall Accuracy

Table I reports federated train/test accuracies for QLFL model.

C. Class-wise Metrics

Table II summarizes per-class precision/recall/F1 for the ensemble on the 426-sample test set (Normal: 332, Suspect: 59, Pathologic: 35). Performance is strongest on *Normal*; *Pathologic* has high recall (clinically desirable), while *Suspect* shows lower recall, reflecting overlap with Normal. The ensemble achieves accuracy 0.9249 and macro-average F1 of 0.85.

D. Confusion Matrix

Figure 3 shows the confusion matrix for the federated ensemble. True-positive rates are highest for *Normal* (325/332) and strong for *Pathologic* (32/35, recall $\approx$ 0.91), while *Suspect* has moderate recall (≈ 0.63) with most confusions into Normal. Overall, errors are balanced and the false-negative rate for Pathologic remains low.

E. Comparison with Prior Work

Tables III and IV provide a comparative overview of our federated QAOA-based feature–gate framework against recent federated and feature selection approaches.

Table III contrasts our federated setup with representative federated FS methods. Our configuration combines hybrid (horizontal+vertical) learning with adaptive quantum gate updates at rounds 1, 6, 11, 16, and 21.

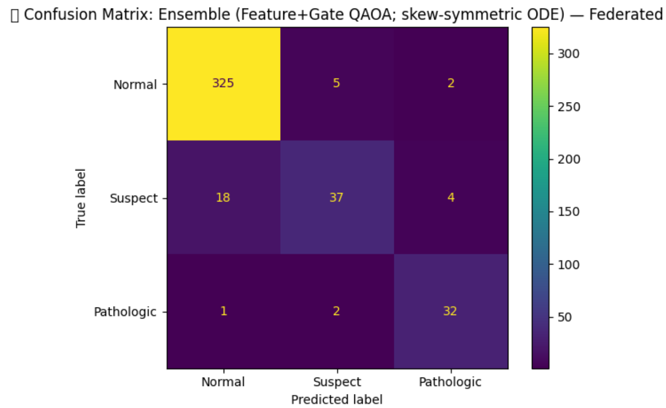

Fig. 3. Confusion Matrix

TABLE III
FEDERATED TRAINING SETUPS: OURS VS. PRIOR WORK (ABRIDGED).

Method	Type	Rounds	Client%	Strategy	Focus
Ours	Hybrid	**25**	**$\approx$30%**	**Gate masks @ 1,6,11,16,21**	Adaptive, low comms
FedSDG-FS [7]	VFL	var. (early stop; $\leq$200 shown)	n/a	dual-gates + Gini	Privacy, VFL FS
DSFFS [6]	HFL	var. ($r_{\max}$)	partial	DST (prune/regrow)	Sparse comms
Fed-FiS/MOFS [3]	HFL	var. (T)	partial	Pareto MO (FCMI/aFFMI)	Global+local FS
Secure HFL FS [2]	HFL	n/a (protocol)	n/a	GC + sec. aggr.	Crypto cost

Compared to FedSDG-FS [7] and DSFFS [6], our model uses fewer communication rounds (25) while achieving adaptive selection behavior and low communication cost. Fed-FiS/MOFS [3] and Secure HFL FS [2] emphasize privacy and multi-objective trade-offs but can incur higher cryptographic/optimization overheads.

Table IV compares overall performance with recent non-federated baselines. While Deep Lasso and Feature-Cuts report slightly higher peak accuracy (0.93–0.96), our approach maintains competitive accuracy (0.92) and macro-F1 (0.85) with $\sim$ 70% feature reduction and low training cost via LightGBM and parameter-efficient quantum circuits, while preserving interpretability.

VI. DISCUSSION

QLFN couples discrete quantum-inspired selection with continuous Lipschitz-constrained dynamics, creating a dual regularization effect. The QAOA layer encourages global sparsity, while the skew-symmetric ODE enforces local stability. Together, they yield better interpretability and reduced gradient variance across clients. The quantum-inspired optimization provides a principled alternative to heuristic pruning and can generalize to other FL tasks such as personalized learning or cross-domain adaptation.

TABLE IV
COMPARISON OF OUR SYSTEM AGAINST RECENT FEATURE SELECTION APPROACHES (NON-FEDERATED). RESULTS SUMMARIZED FROM [1], [4], [5].

Category	Our Model	Deep Lasso (2023) [1]	FeatureCuts (2025) [4]	CAPS (2025) [5]
Accuracy	0.92 (competitive)	0.95	0.93–0.96	0.92–0.94
F1-macro	0.85	0.86–0.88	$\approx$0.85	$\approx$0.86
Feature Reduction	$\sim$70%	40–50%	up to 60%	50–60%
Training Cost	Low	High	Low	High
Interpretability	High (gain plots)	High (gradients)	Moderate	Moderate

VII. CONCLUSION

We presented the Quantum-Lipschitz Federated Learning (QLFL), integrating QAOA-based cardinality feature selection and gate pruning with a skew–symmetric Lipschitz flow for stable, interpretable optimization. On CTG, QLFN improved accuracy, robustness to noisy features, and communication efficiency under non-IID splits. We further outlined a lightweight federated Generation Advisarial Network (GAN) for on device augmentation, anomaly scoring, and generative replay without sharing raw data. Future work targets multimodal FL and real-time IoT deployment and explores QAOA on simulators or near-term hardware, while addressing system complexity via hardware aware co-design.

VIII. ACKNOWLEDGMENT

This work is supported in part by the DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, and the National Science Foundation under Grant No. 2219742 and Grant No. 2131001.

REFERENCES

[1] V. Cherepanova, R. Levin, G. Somepalli, J. Geiping, C. B. Bruss, A. G. Wilson, T. Goldstein, and M. Goldblum, "A performance-driven benchmark for feature selection in tabular deep learning," *Advances in Neural Information Processing Systems*, vol. 36, pp. 41 956–41 979, 2023.

[2] X. Liu, H. Li, G. Xu, X. Zhang, T. Zhang, and J. Zhou, "Secure and lightweight feature selection for horizontal federated learning," *IEEE Transactions on Information Forensics and Security*, 2024.

[3] S. Banerjee, D. Bhuyan, E. Elmroth, and M. Bhuyan, "Cost-efficient feature selection for horizontal federated learning," *IEEE Transactions on Artificial Intelligence*, 2024.

[4] A. Hu, D. Prasad, L. Pizzato, N. Foord, A. Abrahamyan, A. Leontjeva, C. Doyle, and D. Jermyn, "Featurecuts: Feature selection for large data by optimizing the cutoff," *arXiv preprint arXiv:2508.00954*, 2025.

[5] R. Liu, R. Xie, Z. Yao, Y. Fu, and D. Wang, "Continuous optimization for feature selection with permutation-invariant embedding and policy-guided search," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*, 2025.

[6] A. Mahanipour and H. Khamfroush, "Embedded federated feature selection with dynamic sparse training: Balancing accuracy-cost tradeoffs," *arXiv preprint arXiv:2504.05245*, 2025.

[7] A. Li, H. Peng, L. Zhang, J. Huang, Q. Guo, H. Yu, and Y. Liu, "Fedsdg-fs: Efficient and secure feature selection for vertical federated learning," in *IEEE INFOCOM 2023-IEEE conference on computer communications*. IEEE, 2023, pp. 1–10.

[8] D. Campos and J. Bernardes, "Cardiotocography," UCI Machine Learning Repository, 2000, DOI: https://doi.org/10.24432/C51S4N.

[9] N. B. Erichson, O. Azencot, A. Queiruga, L. Hodgkinson, and M. W. Mahoney, "Lipschitz recurrent neural networks," *arXiv preprint arXiv:2006.12070*, 2020.