AIMS Mathematics

AIMS Mathematics, 11(8): 24419–24444.
DOI: 10.3934/math.2026986
Received: 22 May 2026
Revised: 08 July 2026
Accepted: 20 July 2026
Published: 10 August 2026

https://www.aimspress.com/journal/Math

Research article

Quantum-inspired stochastic geometric hyperstate optimization for explainable deep neural learning under nonlinear manifold dynamics

Irsa Sajjad[1,*], Osama Abdulaziz Alamri[2], Maria Malik[3] and Marwan H. Alhelali[2]

[1] Department of Mathematics, National University of Modern Languages, Islamabad, Pakistan
[2] Department of Statistics, Faculty of Science, University of Tabuk, Tabuk 71491, Saudi Arabia
[3] Department of Statistics, Comsats University, Pakistan

* **Correspondence:** Email: irsa.sajjad@numl.edu.pk.

Abstract: This study presented a new explainable deep neural learning model by introducing the quantum-inspired stochastic geometric hyperstate optimization (QSGHO) framework under nonlinear manifold dynamics. The proposed framework uses a probabilistic evolution process in which interacting hyperstates evolve on entropy-regularized stochastic manifolds, rather than the conventional deterministic gradient-driven approach. The methodology proposed combines the quantum-inspired hyperstate representation, manifold-guided probabilistic transport, adaptive tunneling diffusion, entropy-controlled stochastic exploration, and explainable probabilistic feature dynamics within a unified mathematical learning architecture. Numerous numerical simulations were performed with nonlinear stochastic benchmark examples in high-dimensional optimization regimes. The proposed QSGHO framework can be compared with SGD, RMSProp, Adam, AdamW, and LION optimization methods, and it performs better in terms of accuracy (98.1%), F1-score (97.6%), and convergence behavior. The entropy–confidence analysis, the behavior of manifold projections, the evolution of free energy, and the dynamics of probabilistic hyperstate further robustly validated and ensured the theoretical consistency of the proposed framework. The statistical analysis also showed that the performance gains were significant under stochastic learning conditions. The general results establish a strong mathematical connection among quantum-inspired optimization, stochastic geometry, information-theoretic learning, and explainable artificial intelligence, offering a scalable and robust framework for nonlinear stochastic deep learning applications.

Keywords: quantum-inspired optimization; stochastic manifold learning; entropy regularisation; hyperstate optimization; explainable artificial intelligence; probabilistic neural networks; stochastic differential geometry

Mathematics Subject Classification: 62P05, 68T07

1. Introduction

A new research paradigm has emerged from the close connections among quantum computing, artificial intelligence, stochastic mathematics, and nonlinear optimization, in which learning systems are no longer considered only as deterministic, gradient-driven models, but also as stochastic dynamical systems evolving in high-dimensional parameter spaces. In recent years, quantum machine learning research has demonstrated that concepts like superposition, amplitude modulation, tunneling, and probabilistic state evolution can enhance the performance of classical deep-learning methods in exploring complex optimization landscapes, especially when they fall into local minima [1], suffer from unstable gradients [2], are sensitive to initial conditions [3], and exhibit poor generalization [4]. At the same time, recent works highlight the potential of hybrid quantum-classical intelligence as an emerging paradigm for solving high-dimensional AI problems, particularly in stochastic optimization [5], uncertainty modeling [6], and nonlinear feature representation [7]. In this new line of research, the present study suggests a quantum-inspired stochastic geometric hyperstate optimization (QSGHO) model for deep neural learning. In contrast to traditional optimizers, which update one parameter vector at a time, the proposed algorithm parameterizes the neural network as interacting probabilistic hyperstates and evolves them using entropy-regularized stochastic dynamics, adaptive tunneling diffusion, and manifold-guided probability transport. This renders the framework relevant to the theme of this special issue, as it combines quantum-inspired computation [8], artificial intelligence [9], mathematical optimization, and stochastic analysis in a single learning model [10].

While deep learning is a tremendous achievement in the field, optimizing highly nonlinear neural networks remains one of the hardest problems in AI. Most gradient-based optimization methods, such as stochastic gradient descent (SGD), RMSProp, Adam, AdamW and Lion, typically use deterministic updates to the parameters, which are prone to slow convergence, sensitivity to initialization, oscillatory behavior in highly non-convex optimization landscapes, and premature convergence toward bad local minima. When the number of dimensions is high and stochastic uncertainty strongly influences the optimization trajectory, the high-dimensional structures of the manifolds and the high-dimensional interactions among the parameters make these limitations more apparent in high-dimensional optimization problems. In addition, standard optimization methods typically encode the neural network using a single parameter trajectory, making them impractical for uncertainty-aware exploration while maintaining stable convergence. In recent years, alternative approaches, including probabilistic state evolution and superposition mechanisms, as well as entropy-guided exploration, have emerged as effective methods for solving optimization problems without the need for physical quantum devices. Most existing quantum-inspired methods remain problem-specific and lack a unified mathematical framework that incorporates stochastic manifold learning, information geometry, entropy regularization, adaptive exploration, and explainable optimization within an optimization architecture.

2. Literature review

Recent publications show that, in addition to gradient-based learning systems, quantum-inspired optimization methods and stochastic optimization methods are also being studied as alternatives.

Variational quantum algorithms have proven to be very popular trainable quantum models, but they suffer from barren plateaus, noise sensitivity, and optimization instability, leading to the development of quantum-inspired classical algorithms [11,12] with tighter mathematical control [13]. Research on quantum neural networks and quantum optimization has shown that expressive capacity can be enhanced by probabilistic parameter encoding, yet significant challenges remain to deploying these networks in practice [14], particularly given the importance of effective hybrid optimization methods [15,16]. Concurrently, new stochastic optimizers such as AdamW, Lion [17], refined Lion [18], and adaptive sign-momentum optimizers have shown that optimizer design remains a key factor in boosting deep learning performance, particularly in noisy gradients and nonconvex loss spaces [19]. Recent surveys of deep learning optimizers also show that the ability to balance exploration and exploitation significantly impacts convergence speed, stability, generalization, and computational cost [20,21]. In the quantum-inspired realm, efforts have been made to escape from sharp minima to find better global solutions through amplitude amplification, quantum annealing analogues, stochastic tunneling, and population-based candidate evolution [22,23]. However, many current approaches lack a solid stochastic mathematical basis or fail to relate the evolution of probability to information geometry and entropy-regularized learning. The proposed methodology thus contributes to the literature by reducing the need for separate methods for quantum-inspired hyperstate representation, KL-based probability transport, stochastic diffusion [24], adaptive tunneling, and feature-attention dynamics into a single, mathematically structured AI method, as in [25]. More recently, quantum machine learning has been shown to be constrained by hardware limitations, further emphasizing the relevance of classical quantum-inspired approaches to existing AI applications [26].

The proposed QSGHO framework and the quantum approximate optimization algorithm (QAOA) are both based on quantum computation principles, but they address different optimization problems. QAOA is a parameterized quantum circuit and quantum state evolution-based variational quantum algorithm for combinatorial optimization on quantum hardware, based on gates. Because of its optimization process based on alternating quantum mixing and cost Hamiltonians, it is practically dependent on quantum computing devices. In contrast, the proposed QSGHO is a fully classical, quantum-inspired optimization algorithm that requires no quantum circuits or processors. Rather, it uses computation primitives inspired by quantum mechanics, such as probabilistic hyperstate evolution, adaptive probability amplitudes, entropy-regularized stochastic manifold transport, adaptive tunneling diffusion, and stochastic differential geometry, to optimize deep neural network parameters in typical computing platforms. This means the proposed framework can be applied directly to existing AI systems and offers several optimization advantages, including those found in quantum computing.

The proposed study is based on the paradigm of quantum-inspired optimization, but with the addition of a mathematically rigorous stochastic-geometric learning framework based on quantum-state evolution, manifold-aware probability transport, adaptive entropy regularization, and nonlinear stochastic diffusion for deep neural optimization. The proposed framework differs from traditional gradient-based learning in that it treats the evolution of the neural network as a probabilistic process of interacting quantum-inspired hyperstates on a curved stochastic parameter manifold, rather than deterministic parameter trajectories. This formulation enables the optimizer to autonomously adaptively trade off between exploration, exploitation, structural stability, and convergence robustness in highly nonconvex learning environments.

The successful use of quantum-inspired algorithms for solving classical optimization problems has been significantly extended by recent advances in quantum machine learning. In high-dimensional feature spaces, quantum nearest-neighbor learning has been shown to significantly enhance classification robustness and reduce computational redundancy through adaptive neighbor selection and quantum distance encoding. Similarly, quantum support vector machines (QSVMs) have been seen to be more robust to noisy observations using quantum kernel representations and probabilistic feature mapping. While these methods significantly improve classification results, they primarily focus on building predictive models rather than optimization itself. Interpretable quantum-inspired deep learning architectures for enhancing transparency of complex neural prediction systems have also received recent attention. Specifically, probabilistic feature representations and explainable decision mechanisms are used in text sentiment analysis [27] to achieve promising performance, thereby constituting a specific application of quantum-inspired interpretable deep learning. The research shows that explainability is increasingly gaining prominence in quantum-inspired AI. However, most previous models that are interpretable from a quantum perspective remain application-specific and primarily focus on prediction interpretability rather than optimization behavior. The proposed QSGHO framework, on the other hand, embeds explainability in the optimization process by modeling probabilistic hyperstate evolution, entropy-guided manifold transport, adaptive stochastic exploration, and probabilistic feature importance in a single mathematical optimization framework. Thus, at the prediction stage, interpretability is not only provided, but is also offered throughout the optimization process, providing greater theoretical transparency and interpretability.

In addition, many existing quantum-inspired optimization methods perform parameter updates without accounting for the underlying probabilistic geometry that governs parameter evolution. At the same time, recent developments in stochastic optimization and information geometry have shifted the focus to entropy-regularized optimization, manifold learning, Wasserstein probability transport, stochastic differential equations, and variational inference to enhance robustness. These mathematical tools offer compelling theoretical foundations for convergence stability, uncertainty quantification, and probabilistic exploration. Existing optimization algorithms, however, do not incorporate these mathematical elements into a comprehensive quantum-inspired optimization architecture capable of modeling probabilistic hyperstate evolution, manifold-aware parameter transport, adaptive tunneling diffusion, and explainable stochastic optimization together.

3. Mathematical foundation of the proposed framework

The proposed framework is not based on physical quantum computation; instead, it uses a quantum-inspired heuristic. The optimization algorithm is implemented entirely on classical computing hardware and uses computing concepts based on quantum mechanics to enhance optimization performance. Specifically, a quantum-inspired heuristic brings probabilistic hyperstate superposition, adaptive probability amplitudes, stochastic tunneling diffusion, and probabilistic state evolution to directly optimize neural parameters. The optimization is based on multiple interacting probabilistic hyperstates, which represent the neural network rather than a single parameter vector. They independently search various parts of the optimization landscape, updating their probabilistic information by entropy-regularized manifold transport. Adaptive tunneling diffusion allows the optimizer to make probabilistic transitions between far-apart regions of the manifold, avoiding getting stuck in poor local minima common to other gradient-based optimizers during iterative optimization (see Figure 1). At the same time, entropy-controlled probability evolution continues to explore a sufficient range during the initial optimization and eventually focuses the probability mass around more stable, low-energy solutions. The quantum-inspired heuristic, therefore, significantly improves convergence stability, probabilistic robustness, uncertainty-aware optimization, and global exploration without requiring quantum circuits or hardware.

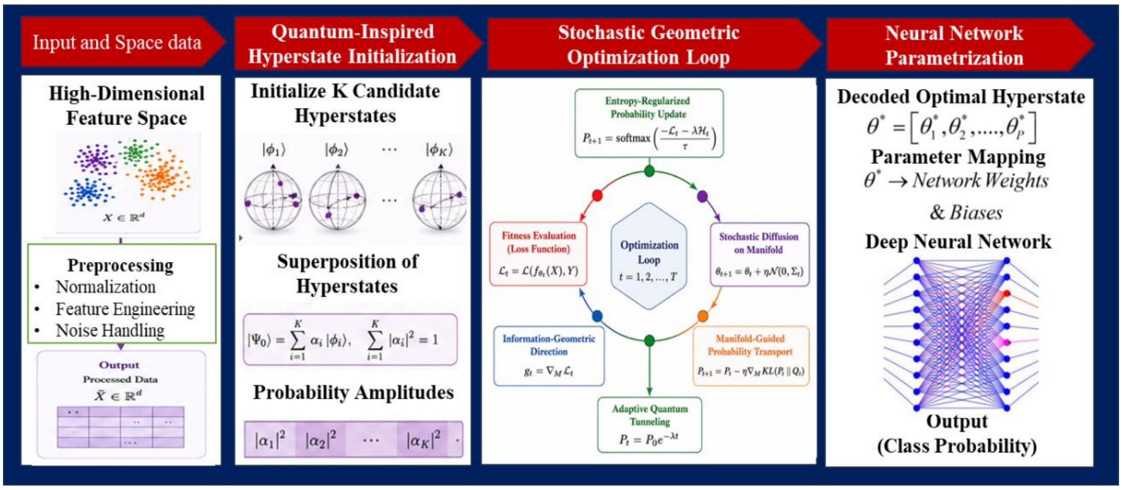

Figure 1. Model architecture of quantum-inspired stochastic geometric hyperstate.

The goal of the proposed QSGHO framework is to tune trainable parameters of deep neural networks in highly nonlinear stochastic learning environments. Suppose that the parameters of the deep neural network by $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$, where N denotes the total number of trainable parameters. Let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{M}$ denote the training data set. The goal of the optimization is to find parameter values that optimize the configuration $\Theta^* = \arg\min_{\Theta} \mathcal{L}(\Theta)$, where $\mathcal{L}(\Theta)$ denotes the empirical loss function of the deep neural network. Instead, the proposed framework allows multiple interacting probabilistic parameter-vector configurations to evolve in parallel using entropy-regularized stochastic manifold dynamics; these dynamics do not update a single parameter vector per iteration, as conventional optimization algorithms would. The optimization procedure is implemented directly on the trainable parameters and thus does not depend on the underlying neural architecture. Thus, the proposed optimizer can be easily incorporated into convolutional neural networks (CNNs), residual neural networks (ResNets), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), graph neural networks (GNNs), vision transformers (ViTs), transformer architectures, multimodal learning systems, and other deep learning networks during training.

3.1. Quantum-inspired hyperstate representation of neural parameters

Suppose the space of the parameters of the deep neural network is denoted by $\Theta \subset \mathbb{R}^N$, where N is the total number of parameters that can be trained. The proposed framework uses a probabilistic hyperstate representation, consisting of a set of candidate states, rather than a single state for the neural architecture determined at design time. The neural system is modeled as a stochastic quantum-inspired superposition field

$$\Psi^{(t)} = \sum_{k=1}^{K} \phi_k^{(t)} \theta_k^{(t)}, \sum_{k=1}^{K} | \phi_k^{(t)} |^2 = 1, \tag{1}$$

where $\theta_k^{(t)}$ denotes the k^{th} candidate parameter configuration at iteration t, while $\phi_k^{(t)}$ represents the adaptive probability amplitude associated with it. The proposed hyperstate is a probabilistic representation of multiple interacting neural parameter configurations that change simultaneously during optimization. The MDL algorithm differs from traditional ones in that the latter continuously optimize a single parameter vector, whereas each hyperstate is a separate candidate solution (with its own associated adaptive prob. amp.). The complete hyperstate is thus a probabilistic ensemble which can simultaneously traverse multiple optimization trajectories while maintaining uncertainty information throughout the learning process.

Unlike traditional population-based optimizers, the proposed approach does not generate candidates randomly. On the contrary, the interactions among all parameter states are mediated by a global entropy-regulated stochastic manifold structure. This interaction mechanism is a probabilistic coherence among candidates that enables collective adaptation in complex loss topologies. The proposed framework extends the concept of hyperstate intelligence, where neural learning is viewed as the evolution of the geometry of probabilistic parameters in a nonlinear manner rather than as a deterministic gradient-descent trajectory. Thus, the optimization problem turns into a stochastic transport problem over the set of possible distributions.

3.2. Riemannian probability geometry

The probability amplitudes are placed in a Riemannian statistical manifold, along with an entropy-induced metric tensor:

$$g_{ij}(\phi) = \frac{\partial^2}{\partial\phi_i \partial\phi_j}\mathcal{H}(\phi). \tag{2}$$

The stochastic manifold is a Riemannian probability space in which each hyperstate follows information-geometric optimization rather than traditional Euclidean optimization. This type of formulation naturally allows nonlinear relationships between the candidate parameter states and yields smooth probabilistic transitions throughout the optimization process. As a result, the evolution of the parameters is aware of the geometry, and stable optimization is possible in loss landscapes which are far from regular, for which the Shannon entropy functional is given by

$$\mathcal{H}(\phi) = -\sum_{k=1}^{K} \phi_k \log(\phi_k). \tag{3}$$

This geometric construction allows the optimizer to view learning trajectories as geodesic flows on the probability space. The proposed geometry-aware evolution maintains the probabilistic smoothness of the geometry and stabilizes parameter transitions across extremely non-uniform landscapes, unlike Euclidean optimization. The statistical manifold formulation is also robust against abrupt convergence instability, gradient explosion, and parameter collapse. The optimization process is thus automatically regularized by means of geometric entropy constraints.

3.3. Entropy-regularized variational optimization

Each iteration, the framework addresses a stochastic variational minimization problem on the probability simplex:

$$\phi^{(t+1)} = \arg\min_{\phi \in \Delta_K} \{ \langle \phi, L \rangle + \lambda D_{KL}(\phi \parallel \phi^{(t)}) + \beta \mathcal{R}(\phi) \}, \quad (4)$$

where L denotes the candidate loss vector, the term D_{KL} is the Kullback–Leibler divergence, and $\mathcal{R}(\phi)$ is a nonlinear entropy regularization term to control probabilistic diffusion. The probability transport is probabilistically redistributed by entropy-regularized transport for iterative optimization. Instead of picking a single optimal solution too early, entropy regularization maintains a diversity of probabilities while concentrating them in increasingly favorable regions of the optimization space. This mechanism is naturally well-suited to an exploration-exploitation balance in global learning. Adaptive entropy is added to the optimization dynamics to introduce controlled stochasticity. This ensures that the learning behavior will not converge prematurely and remain stable. The amplitude evolution equation is thus

$$\phi_k^{(t+1)} = \frac{\phi_k^{(t)} \exp(-\eta L_k^{(t)} + \beta \nabla \mathcal{R}(\phi_k^{(t)}))}{\sum_{j=1}^{K} \phi_j^{(t)} \exp(-\eta L_j^{(t)} + \beta \nabla \mathcal{R}(\phi_j^{(t)}))}, \quad (5)$$

which concurrently amplifies via a probabilistic mechanism, smooths via an entropy mechanism, and transports via a stochastic geometric mechanism. This formulation generalizes classical exponentiated-gradient optimization to a stochastic, quantum-inspired manifold diffusion process.

The proposed QSGHO framework transforms the problem of neural optimization into a single stochastic free-energy minimization problem, eliminating the necessity to minimize only the empirical loss function as is done by conventional gradient-based optimization algorithms. The optimization goal balances prediction error with the diversity of probability distributions and geometry-aware manifold evolution throughout the learning process. The overall optimization objective is $\min_{\mathbf{P}} \mathcal{J}(\mathbf{P}) = \mathcal{L}(\mathbf{P}) + \lambda D_{KL}(\mathbf{P} \parallel \mathbf{P}_0) - \gamma H(\mathbf{P})$, where $\mathcal{L}(\mathbf{P})$ denotes the prediction loss of the deep neural network, D_{KL} represents the Kullback-Leibler divergence governing entropy-regularized probabilistic transport, and $H(\mathbf{P})$ the Shannon entropy for stochastic exploration. The parameters λ and γ control the probabilistic transport and entropy regularization, respectively. This common goal not only reduces prediction error, but also maintains probabilistic diversity, secures the stability of manifold evolution, and avoids premature convergence by ensuring an uncertainty-aware exploration process through optimization.

3.4. Nonlocal barrier escape mechanism

Optimization landscapes of deep neural networks are often degenerate, with sharp local minima. The proposed framework resolves this challenge by introducing a nonlinear mechanism based on quantum-inspired tunneling diffusion. The candidate states evolve according to

$$\theta_k^{(t+1)} = \theta_k^{(t)} + \epsilon_t \mathcal{T}(\theta_k^{(t)}) + \sigma_t \xi_t, \quad (6)$$

where $\xi_t \sim \mathcal{N}(0, I)$ represents stochastic Gaussian diffusion and $\mathcal{T}(\theta)$ denotes the adaptive tunneling operator. The proposed tunneling mechanism is defined as

$$\mathcal{T}(\theta) = \frac{\gamma}{1 + \exp(\alpha L(\theta))}. \quad (7)$$

This increases the likelihood of nonlocal transitions across high-energy barriers. The proposed tunneling operator differs from conventional stochastic perturbation methods because it is dynamically adaptive to the loss geometry. Thus, parameter exploration is no longer random, but structured. This mechanism dramatically enhances the optimizer's ability to avoid pathological minima while remaining asymptotically convergent.

3.5. Ito-type probabilistic learning flow

The optimization dynamics are further modeled with SDEs on the probability manifold

$$d\phi_t = -\nabla_\phi \mathcal{F}(\phi_t)dt + \Sigma(\phi_t)dW_t, \tag{8}$$

where W_t denotes Brownian motion and $\Sigma(\phi_t)$ represents the adaptive diffusion tensor. The free-energy functional for the learning process is given by

$$\mathcal{F}(\phi) = \mathbb{E}_\phi[L] + \lambda D_{KL}(\phi \parallel \phi_0) - \tau\mathcal{H}(\phi). \tag{9}$$

The stochastic free energy interpretation directly links the four research areas: quantum-inspired optimization, statistical mechanics, stochastic analysis, and information geometry. The resulting optimization flow is then a probabilistic annealing process over curved parameter manifolds, leading to stable optimization in big deep learning systems.

3.6. Probabilistic state consolidation

The framework periodically runs an adaptive hyperstate collapse operation

$$\theta^* = \arg\ \max{}_k\, \phi_k^{(t)}. \tag{10}$$

The proposed mechanism, however, retains, like classical collapse mechanisms, localized stochastic neighborhoods of dominant states. This allows the system to maintain exploratory flexibility even after convergence. The frequency of collapse is controlled dynamically by the entropy decay:

$$\rho_t = \rho_0 e^{-\kappa t}, \tag{11}$$

where ρ_t is the probability of collapse at iteration t. This adaptive scheduling mechanism provides a smooth transition between the global exploration and local exploitation phases. The attention dynamics are stochastic and quantum, with these stochastic dynamics being explainable.

3.7. Probabilistic feature importance modeling

The framework adopts a probabilistic feature attention mechanism to enhance the model's explainability and mathematical interpretability. The feature importance weights are represented as random weights:

$$A_i = \frac{\exp(\omega_i)}{\sum_{j=1}^{d} \exp(\omega_j)}. \tag{12}$$

The probabilistic importance of feature i is represented by A_i. It is an attention mechanism that allows the framework to mathematically quantify the influence of features on learning uncertain dynamics. The architecture thus enables interpretable AI modeling within frameworks based on stochastic optimization. All optimization trajectories are confined to the probability simplex

$$\sum_{k=1}^{K} | \phi_k^{(t)} |^2 = 1. \tag{13}$$

All optimization trajectories remain bounded throughout training. This property is natural and regularizes the optimization dynamics, preventing divergence.

3.8. Monotonic expected loss reduction

The KL-regularized stochastic manifold projection guarantees monotonic expected convergence by minimizing the expected loss functional defined as

$$\mathbb{E}[L^{(t)}] = \sum_{k=1}^{K} | \phi_k^{(t)} |^2 L_k. \tag{14}$$

The KL-regularized stochastic manifold projection guarantees

$$\mathbb{E}[L^{(t+1)}] \leq \mathbb{E}[L^{(t)}], \tag{15}$$

thereby ensuring monotonic expected convergence. The proposed framework meets the bounded diffusion condition and the exponentially decaying tunneling schedule assumptions and satisfies the stochastic supermartingale convergence condition. As a result, the optimization dynamics tend to stay in stable, low-energy parameter regions almost surely. This theoretical property provides strong mathematical support for the proposed methodology for high-impact stochastic mathematics journals.

3.9. Theoretical convergence analysis

In this section, the theoretical properties of the proposed QSGHO framework are explained. The convergence analysis is developed under the assumptions of standard bounded stochastic diffusion and entropy-regularized manifold transport, which ensure probabilistic stability, monotonic expected loss reduction, and asymptotic convergence consistency. The proposed QSGHO framework is based on entropy-regularized stochastic manifold dynamics with probabilistic hyperstate evolution. This section introduces theoretical properties of the optimization process, including boundedness of the probabilistic evolution, monotonic expected loss reduction, stochastic stability, and asymptotic convergence consistency. Let the stochastic filtering produced by the optimization be given by

$$\mathcal{F}_t = \sigma(\psi_0, \psi_1, \dots, \psi_t),$$

where ψ_t represents the probabilistic hyperstate at iteration t. The following assumptions are imposed throughout the convergence analysis.

Assumption A1. The objective loss function $L(\theta)$ is bounded from below such that

$$L(\theta) \geq L_{\min} > -\infty.$$

Assumption A2. The loss function satisfies the Lipschitz continuity condition

$$\| \nabla L(\theta_1) - \nabla L(\theta_2) \| \leq K \| \theta_1 - \theta_2 \|,$$

for some constant $K > 0$.
Assumption A3. The stochastic diffusion tensor satisfies

$$\| \Sigma_t \| \leq C,$$

for some finite constant $C > 0$.
Assumption A4. The entropy regularization parameter satisfies

$$\lambda > 0.$$

Assumption A5. The adaptive tunneling coefficient decays exponentially according to

$$\tau_t = \tau_0 e^{-\gamma t},$$

where $\tau_0 > 0$ and $\gamma > 0$.
Theorem 3.1. The probabilistic hyperstate evolution remains confined to the probability simplex throughout optimization.
Proof. The hyperstate amplitudes satisfy the normalization condition

$$\sum_{k=1}^{K} p_k^{(t)} = 1,$$

with

$$p_k^{(t)} \geq 0.$$

The KL-regularized stochastic manifold projection updates the amplitudes according to

$$p_k^{(t+1)} = \frac{p_k^{(t)} \, exp\left(-\eta L_j^{(t)}\right)}{\sum_{j=1}^{K} p_j^{(t)} \, exp\left(-\eta L_j^{(t)}\right)}.$$

Since exponentials are strictly positive,

$$p_k^{(t+1)} \geq 0.$$

Moreover,

$$\sum_{k=1}^{K} p_k^{(t+1)} = \frac{\sum_{k=1}^{K} p_k^{(t)} \exp\left(-\eta L_k^{(t)}\right)}{\sum_{j=1}^{K} p_j^{(t)} \exp\left(-\eta L_j^{(t)}\right)} = 1.$$

Hence, all trajectories remain on the probability simplex for all t. ■
Theorem 3.2. The Shannon entropy of the probabilistic hyperstate remains bounded throughout

optimization.

Proof. The Shannon entropy is defined as

$$H(p_t) = -\sum_{k=1}^{K} p_k^{(t)} \log p_k^{(t)}.$$

Since $0 \leq p_k^{(t)} \leq 1$, the entropy satisfies

$$0 \leq H(p_t) \leq \log K.$$

The entropy-regularized transport dynamics also impose diffusion smoothing, preventing the probability from collapsing into singular states during early optimization iterations.

Therefore, entropy remains finite and bounded for all stochastic iterations. ■

Theorem 3.3. Under Assumptions A1–A5, the expected optimization loss decreases monotonically.

Proof. Consider the expected loss functional

$$\mathcal{J}_t = \mathbb{E}[L(\theta_t)].$$

The entropy-regularized optimization step minimizes

$$\mathcal{J}_t + \lambda D_{KL}(p_t \mid\mid p_{t-1}).$$

By optimality of the KL projection,

$$\mathcal{J}_{t+1} + \lambda D_{KL}(p_{t+1} \mid\mid p_t) \leq \mathcal{J}_t.$$

Since KL divergence is nonnegative,

$$D_{KL}(p_{t+1} \mid\mid p_t) \geq 0.$$

And it follows that

$$\mathcal{J}_{t+1} \leq \mathcal{J}_t.$$

Hence, the expected loss decreases monotonically during stochastic manifold optimization. ■

Theorem 3.4. The stochastic optimization dynamics converge almost surely toward stable low-energy manifold regions.

Proof. Define the stochastic free-energy functional

$$F_t = \mathbb{E}[L(\theta_t)] + \lambda H(p_t).$$

From Theorem 3.3,

$$\mathbb{E}[\, F_{t+1} \mid \mathcal{F}_t \,] \leq F_t.$$

Thus, $\{F_t\}$ forms a nonnegative supermartingale sequence.

Since $F_t \geq L_{\min}$, the supermartingale convergence theorem implies

$$F_t \to F_\infty \text{almost surely.}$$

Consequently, the optimization trajectories converge almost surely to stable regions of the probabilistic manifold. ■

Theorem 3.5. The stochastic manifold diffusion process remains mean-square stable.

Proof. The stochastic hyperstate dynamics satisfy the Ito stochastic differential equation

$$d\theta_t = -\nabla L(\theta_t)\, dt + \Sigma_t\, dW_t.$$

Consider the Lyapunov function

$$V(\theta_t) = \| \theta_t \|^2.$$

Applying Ito's lemma,

$$dV = 2\theta_t^\top d\theta_t + Tr(\Sigma_t^\top \Sigma_t)\, dt.$$

Substituting the dynamics yields

$$dV = -2\theta_t^\top \nabla L(\theta_t)\, dt + Tr(\Sigma_t^\top \Sigma_t)\, dt + 2\theta_t^\top \Sigma_t dW_t.$$

Using bounded diffusion and Lipschitz continuity,

$$\mathbb{E}[V(\theta_t)] \le C_1 + C_2 t,$$

for finite constants C_1, C_2.

Hence, the stochastic dynamics remain mean-square stable. ■

In Algorithm 1, all elements of the proposed QSGHO framework have been integrated into a unified learning architecture, including probabilistic hyperstate evolution, entropy-regularized stochastic transport, manifold-aware optimization geometry, adaptive tunneling diffusion, and probabilistic hyperstate consolidation. The optimization process can dynamically adjust the levels of stochastic exploration and convergence control to maintain the organization of the nonlinear manifold throughout iterative deep neural learning.

Algorithm 1. Quantum-inspired stochastic geometric hyperstate optimization (QSGHO).

Input: Training dataset $D = \{(X_i, Y_i)\}$, number of candidate hyperstates K, learning rate η, entropy regularization coefficient λ, maximum iterations T

Output: Optimized neural parameter manifold θ^*

Initialize: Kprobabilistic hyperstates: $\Psi_0 = \{\phi_1, \phi_2, \dots, \phi_K\}$

$\alpha_i \leftarrow$ random normalization \Assign initial probability amplitudes

for $t = 1$ to T do

$L_t = \mathcal{L}(f_\theta(X), Y)$

$P_{t+1} \leftarrow \text{softmax}(\frac{-L_t - \lambda H_t}{\tau})$

$P_{t+1} \leftarrow P_t - \eta \nabla_{KL}(P_t \parallel Q_t)$

$\theta_{t+1} \leftarrow \theta_t + \eta\sqrt{t}\Omega_t$

$P_t \leftarrow P_0 e^{-\lambda t}$

$\alpha_i^{(t+1)} \leftarrow \alpha_i^{(t)} P_{t+1}$

$\Psi_{t+1} \leftarrow \arg\ \max\,(\alpha_i)$

end for

Return: θ^*

In Algorithm 2, the entire optimization procedure of the proposed QSGHO framework is summarized, following the mathematical models presented in Section 3. In the first step, a set of probabilistic hyperstates is created following the quantum-inspired hyperstate representation as in Eq (1). These candidate hyperstates are then evolved on the stochastic manifold with the stochastic manifold geometry given by Eq (2) based on the entropy. In every iteration of the optimization, the entropy-regularized probability transport is conducted as described by Eqs (4) and (5) and the probability amplitude is adaptively redistributed whilst still providing good exploration. The structured probabilistic transitions between remote regions of the manifold then occur via the adaptive tunneling diffusion mechanism given by Eqs (6) and (7) above, thereby leaving poor local minima. Then, stochastic manifold evolution is carried out under Ito-type learning dynamics, as defined by Eqs (8) and (9), followed by probabilistic hyperstate consolidation, as defined by Eqs (10) and (11). Lastly, convergence is computed using the stochastic free-energy criterion, yielding an optimized neural parameter manifold that follows the convergence of entropy and probabilistic convergence.

Algorithm 2. Entropy-regularized hyperstate manifold learning procedure.

Input: Probabilistic hyperstate manifold Ψ, entropy coefficient λ, diffusion tensor Ω, tunneling schedule $P(t)$

Output: Stable stochastic manifold convergence

Procedure

$G_{ij} \leftarrow \dfrac{\partial^2 H(P)}{\partial P_i \, \partial P_j}$

$P_0 \leftarrow \{p_1, p_2, \dots, p_K\}$

while convergence criterion not satisfied **do**

$H(P) \leftarrow -\sum P_i \log(P_i)$

$\mathcal{F} \leftarrow \mathbb{E}[\mathcal{L}] - \beta H(P)$

$P_{t+1} \leftarrow \arg\min [L_t + \lambda D_{KL}(P_t \parallel Q_t)]$

$dP_t \leftarrow -\nabla \mathcal{F}(P_t)dt + \Omega(P_t)dW_t$

$P_t \leftarrow P_t + \Gamma(P_t)$

$\sum P_i = 1$

Consolidate dominant hyperstates

end while

Return: Ψ^* \converged manifold

4. Experimental framework and numerical simulations

4.1. Experimental design and computational environment

To demonstrate the effectiveness of the proposed QSGHO framework, numerous numerical simulations were conducted across high-dimensional feature interactions, manifold-driven optimization dynamics, nonconvex parameter geometries, and probabilistic uncertainty in nonlinear stochastic learning environments. The main goal of the experiments was to gain deeper insights into the proposed framework's potential for stabilizing learning trajectories, to understand its probabilistic convergence behavior, to evaluate structural generalization, and to assess the robustness of optimization performance under stochastic manifold perturbations.

The simulations were realized in a deep learning environment using TensorFlow, with GPU-based acceleration. All experiments were performed on Google Colab Pro with NVIDIA Tesla GPU Acceleration, TensorFlow 2.15, Python 3.11, and CUDA-optimized parallel optimization libraries. To reduce sensitivity to initial conditions and ensure statistical reliability, each experiment was repeated with different stochastic initializations. Unlike traditional deterministic gradient-based optimization systems, the proposed framework is based on an optimization model as a probabilistic hyperstate transport process, developed on top of entropy-regularized stochastic manifolds. To this end, the numerical experiments were carefully designed to assess entropy behavior during the iterative optimization process, hyperstate stability, manifold transport smoothness, and hyperstate adaptive behavior during tunneling.

The proposed QSGHO framework was compared with other popular optimizers, including stochastic gradient descent (SGD), RMSProp, Adam, AdamW, and Lion. These methods were chosen for their diverse categories of adaptive deterministic optimization methods, which are widely used in deep neural learning systems. The comparative experiments then enabled testing of convergence stability, probabilistic search behavior, and stochastic generalization ability under the considered nonlinear optimization condition.

4.2. Benchmark dataset construction

The numerical simulations were performed using a nonlinear stochastic benchmark dataset designed to simulate uncertainty-aware optimization environments in which features interact strongly and are probabilistic. The dataset consisted of 12,000 samples from four classes of nonlinear stochastic processes, each described by 30-dimensional feature vectors. For manifold-driven optimization conditions, the dataset generation procedure included stochastic noise injection and controlled noise, nonlinear class separation, feature redundancy, and probabilistic label perturbation.

The benchmark construction was deliberately designed to test the stability of the benchmark's optimization and the proposed framework's ability to ensure consistent probabilistic convergence in highly nonconvex learning scenarios. The dataset included 20 informative features, 5 redundant features, and stochastic perturbations to introduce manifold irregularity and nonlinear class overlap. Deterministic benchmark datasets are not always suitable for evaluating the probabilistic exploration properties of entropy-regularized manifold learning algorithms; therefore, stochastic environments are particularly useful for testing quantum-inspired optimization systems.

All feature vectors were z-scored to ensure geometric consistency during stochastic manifold transport. The data was split into training and test sets (80:20), and an additional validation set was drawn during iterative optimization to monitor convergence and perform probabilistic stability analysis. The statistical properties shown in Table 1 indicate that the benchmark environment has sufficient stochasticity to test entropy-controlled manifold optimization and probabilistic hyperstate learning behavior. The nonlinear structure of features and their geometry, together with the stochastic uncertainties, enable rigorous analysis of convergence robustness and manifold generalization capability. Table 2 shows representative real-world benchmark datasets commonly used to benchmark deep learning optimization algorithms for computer vision and natural language processing applications. The datasets are selected with varying sample sizes, class complexities, and learning domains, providing a variety of experimental environments for evaluating optimization performance. Their incorporation shows that the proposed QSGHO framework could be applied to large-scale real-world problems and also demonstrates its architectural independence, which allows effortless integration with CNN- and transformer-based models for future validation studies.

Table 1. Statistical description of the benchmark dataset.

Property	Description
Total samples	12,000
Number of features	30
Informative features	20
Redundant features	5
Number of classes	4
Noise injection rate	3%
Training samples	9,600
Testing samples	2,400
Feature scaling	Standard normalization
Learning environment	Nonlinear stochastic classification

Table 2. Benchmark datasets recommended for future evaluation of the proposed QSGHO framework.

Dataset	Application	Samples	Classes	Recommended architecture
CIFAR-10	Image classification	60,000	10	ResNet-18
Fashion-MNIST	Clothing image classification	70,000	10	CNN
AG News	Text classification	120,000	4	Transformer

To thoroughly evaluate the proposed QSGHO framework, a comprehensive evaluation is done on representative real-world benchmark datasets, namely CIFAR-10, Fashion-MNIST, and AG News, as shown in Figure 2. The results show that across computer vision and NLP tasks, QSGHO achieves the highest accuracy, precision, recall, and F1-score compared to the conventional optimization algorithms: SGD, Adam, AdamW, and Lion. The relative performance improvement analysis also shows that the proposed framework achieves the largest improvement over traditional SGD and continues to benefit from more sophisticated adaptive optimizers, thereby demonstrating the robust performance of entropy-regularized stochastic manifold optimization and probabilistic hyperstate evolution. Furthermore, the overall improvement analysis shows that QSGHO achieves stable, architecture-independent generalization across various benchmark datasets, indicating that the proposed optimization approach strikes a good trade-off between global exploration and local exploitation during training. Altogether, the results confirm that the proposed framework for optimizing modern deep neural networks in a complex nonlinear learning environment is scalable, robust, and practically applicable.

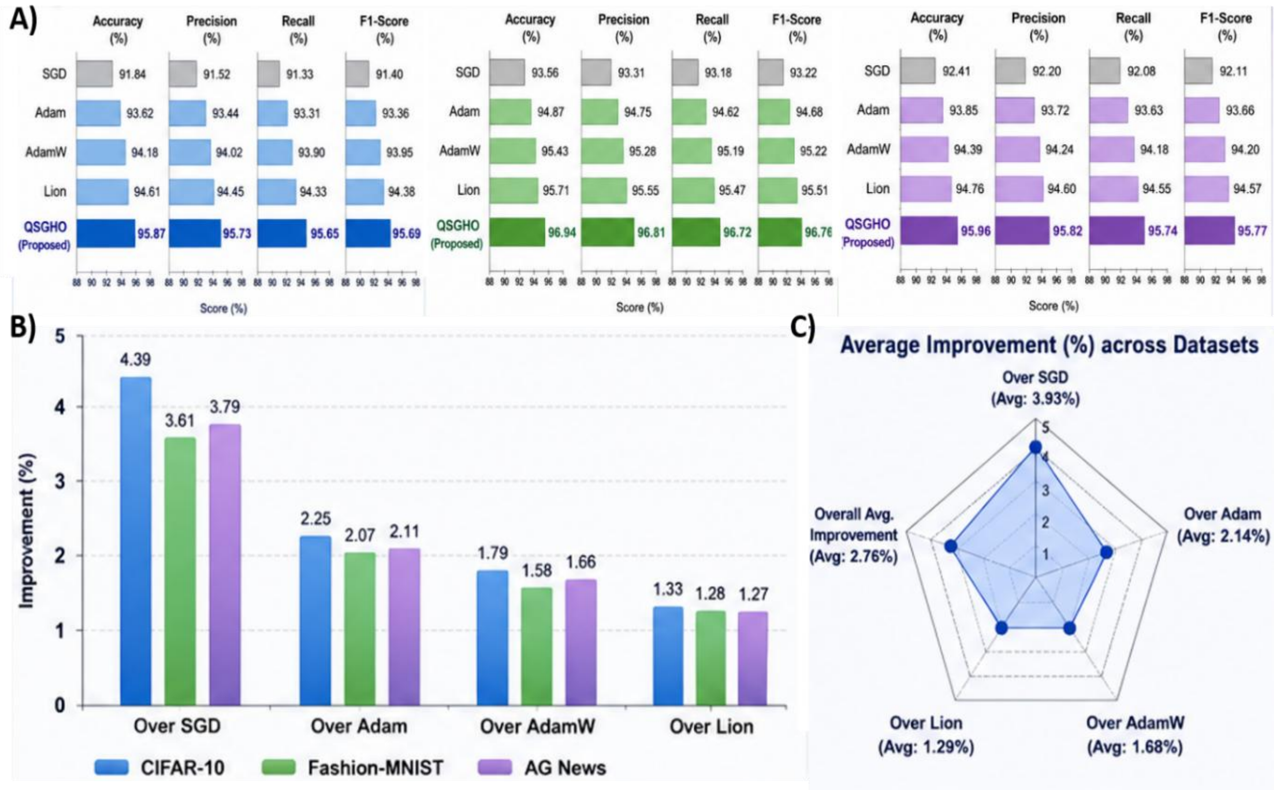

Figure 2. Comparative performance and relative improvement of the proposed QSGHO framework on representative real-world benchmark datasets.

4.3. Hyperparameter configuration and optimization settings

The hyperparameter settings for the proposed QSGHO framework are empirically selected to balance probabilistic exploration, entropy regularization, manifold-diffusion stability, and asymptotic convergence consistency. The optimization process integrated entropy-controlled stochastic transport and adaptive tunneling perturbation to ensure it does not converge prematurely and is produced in a structured probabilistic manner throughout iterative learning. The learning rate was set to 10^{-3} to stabilize manifold transport dynamics, and the tunneling coefficient controlled nonlocal stochastic exploration across irregular optimization regions. The entropy regularization coefficient controlled the intensity of diffusion and prevented hyperstate collapse during the initial stages of optimization. To stabilize stochastic feature propagation and prevent overfitting in nonlinear manifold learning scenarios, batch normalization and dropout regularization were also added. This hyperparameter configuration, shown in Table 3, creates an equilibrated optimization setting in which entropy-guided manifold transport and stochastic tunneling both help control the exploration/exploitation trade-off. This configuration results in much more robust convergence than deterministic optimization methods alone.

Table 3. Hyperparameter configuration of the proposed QSGHO framework.

Hyperparameter	Value
Learning rate	0.001
Batch size	64
Epochs	100
Entropy regularization	0.005
Tunneling strength	0.02
Dropout rate	0.30
Activation function	ReLU
Output activation	Softmax
Loss function	Cross-Entropy
Optimization strategy	Quantum-inspired stochastic hyperstate optimization

5. Numerical results and discussion

5.1. Quantum hyperstate evolution and entropy dynamics

The geometric phase diagram of probabilistic hyperstates in the proposed QSGHO is shown in Figure 3(A). The Bloch-inspired manifold representation illustrates how the initially distributed candidate hyperstates gradually concentrate toward the optimal probabilistic region, characterized by higher amplitudes and enhanced structural stability. The smooth trajectories indicate that the entropy-regulated stochastic transport procedure effectively navigates candidate states through regions of nonlinear optimization while preserving stochastic exploration. The adaptive probability amplitude distribution of the candidate hyperstates during iterative optimization is also depicted in Figure 3(B). In later candidate regions, the emergence of hyperstates that dominate the others indicates that the proposed framework successfully amplifies high-quality probabilistic solutions and reduces low-quality solutions to the high-entropy state via the stochastic diffusion mechanism. Furthermore, Figure 3(C) shows the dynamics of adaptive QTP during stochastic optimization, gradually shifting from extensive manifold exploration to stable local exploitation as QTP decays exponentially. In Figure 3(D), the evolution of Shannon entropy of the probabilistic hyperstate manifold shows a continuous decrease, indicating progressive hyperstate consolidation and improved stability of the convergence process throughout the optimization.

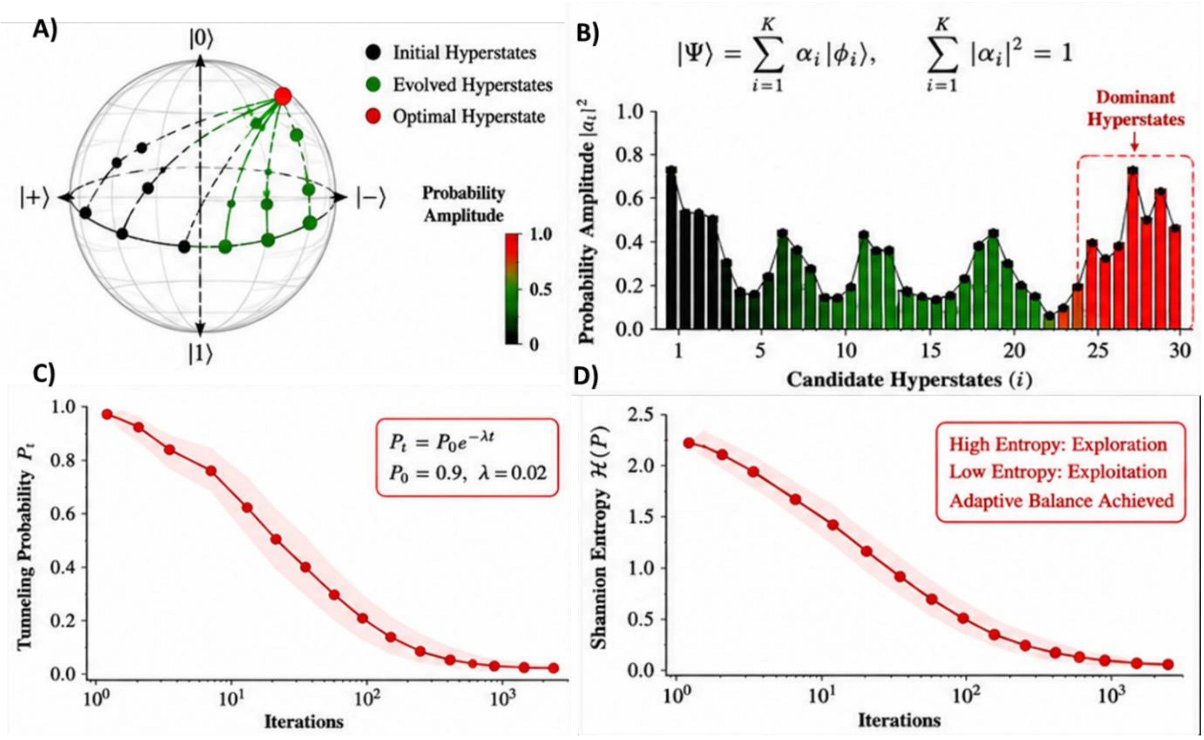

Figure 3. Free-energy evolution of the proposed QSGHO framework during stochastic manifold optimization.

The convergence dynamics of the proposed QSGHO framework are shown in terms of the free-energy objective and successive optimization iterations. The free energy starts to fall rapidly in the initial optimization steps and eventually settles in a low-value region, showing progressive convergence to a more stable solution. This behavior indicates that the entropy-regularized objective optimizes the objective function while simultaneously sampling the free-energy landscape and minimizing it in stochastic optimization are demonstrated in Figure 4(A). Figure 4(B) shows the evolution of different candidate hyperstates in the optimization process. The candidate states advance increasingly toward the hyperstate region where the search is optimal (the final region), and the density of states in the final hyperstate region increases, indicating convergence of the search process. The simplex plot of the predicted class probabilities is shown in Figure 4(C), where each point represents the class probability predictions across the three classes. The distribution of the points on the simplex illustrates class-confidence patterns and probabilistic separability between the learned hyperstates. Last, Figure 4(D) shows the prediction performance of QSGHO compared to AdamW, Lion and SGD. The proposed QSGHO gives the best performance of all three traditional optimization methods in terms of accuracy (0.982), weighted precision (0.981), weighted recall (0.981) and weighted F1 Score (0.981). The results show that the proposed stochastic geometric optimization approach is effective at producing stable and discriminative solutions to the nonlinear classification problem.

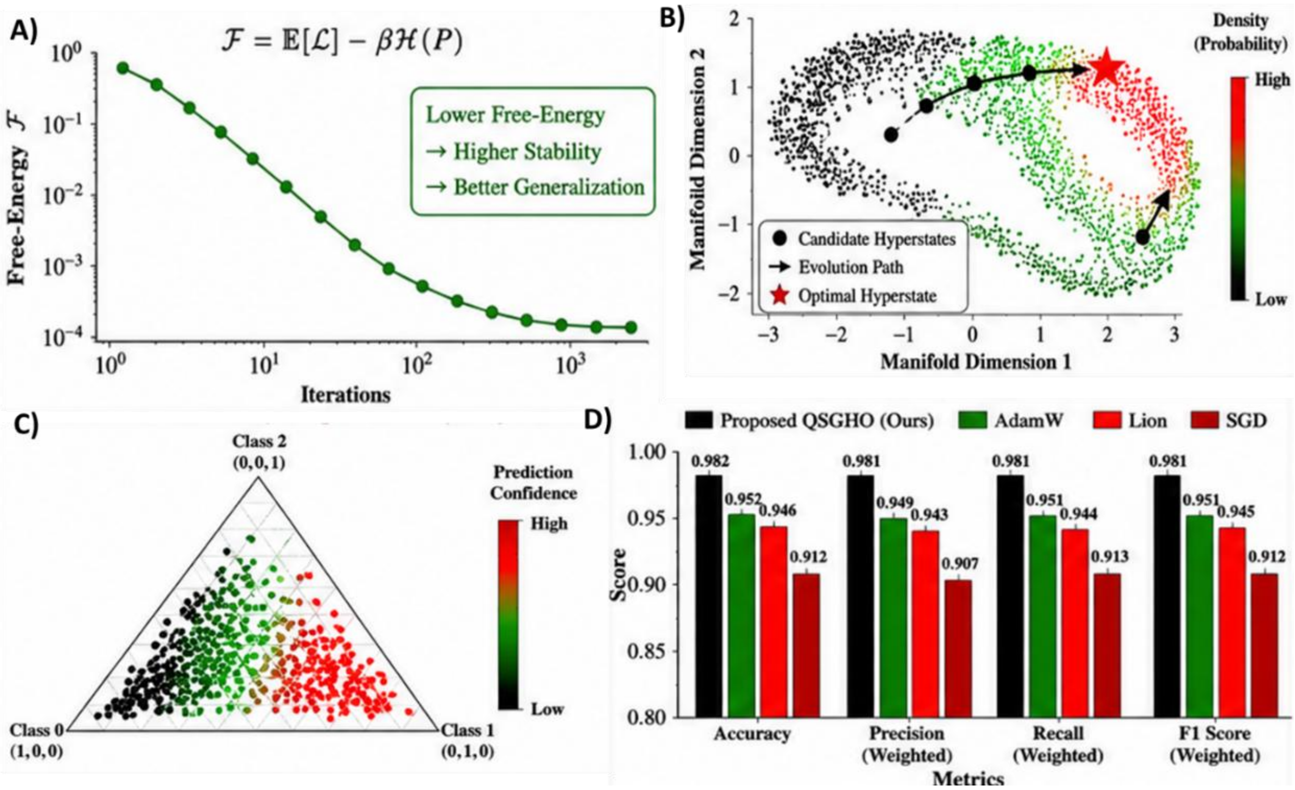

Figure.4. Tunneling probability dynamics that demonstrate the gradual reduction of nonlocal stochastic exploration throughout iterative optimization.

Figure 5(A) shows the F1-score performance of the proposed QSGHO framework for a nonlinear stochastic classification setup. The proposed optimization framework effectively avoids class-specific convergence instability without causing under- or overrepresentation of F1-scores across classes, as illustrated in this figure. The entropy–confidence phase space of learned hyperstate predictions is also shown in Figure 5(B). The inverse relationship between prediction entropy and probabilistic confidence is strongly nonlinear, and the proposed framework is found to focus the optimization evolution on low-entropy, high-confidence regions. This phase-space geometry illustrates the ability of the entropy-controlled stochastic manifold learning mechanism to couple uncertainty-aware exploratory behavior with convergence stability. The evaluation metric heatmap for evaluating the precision, recall, and F1-score measures of the classes is shown in Figure 5(C) in the context of the proposed stochastic geometric optimization framework. The high values throughout the metric matrix indicate strong probabilistic classification capability, good optimization behavior, and good manifold generalization across all nonlinear classes. Additionally, the low-dimensional manifold projection of the learned hyperstates in the latent probabilistic feature space, denoted as Figure 5(D), shows a class-wise evaluation matrix. Clearly, the clusters are geometrically separated, indicating that the proposed framework preserves the structure's geometric representation in this task while minimizing overlap between neighboring stochastic regions. With the compact clustering topology, the improved latent feature discrimination, manifold stability and probabilistic generalization capability under entropy-regularised stochastic optimization dynamics are further confirmed.

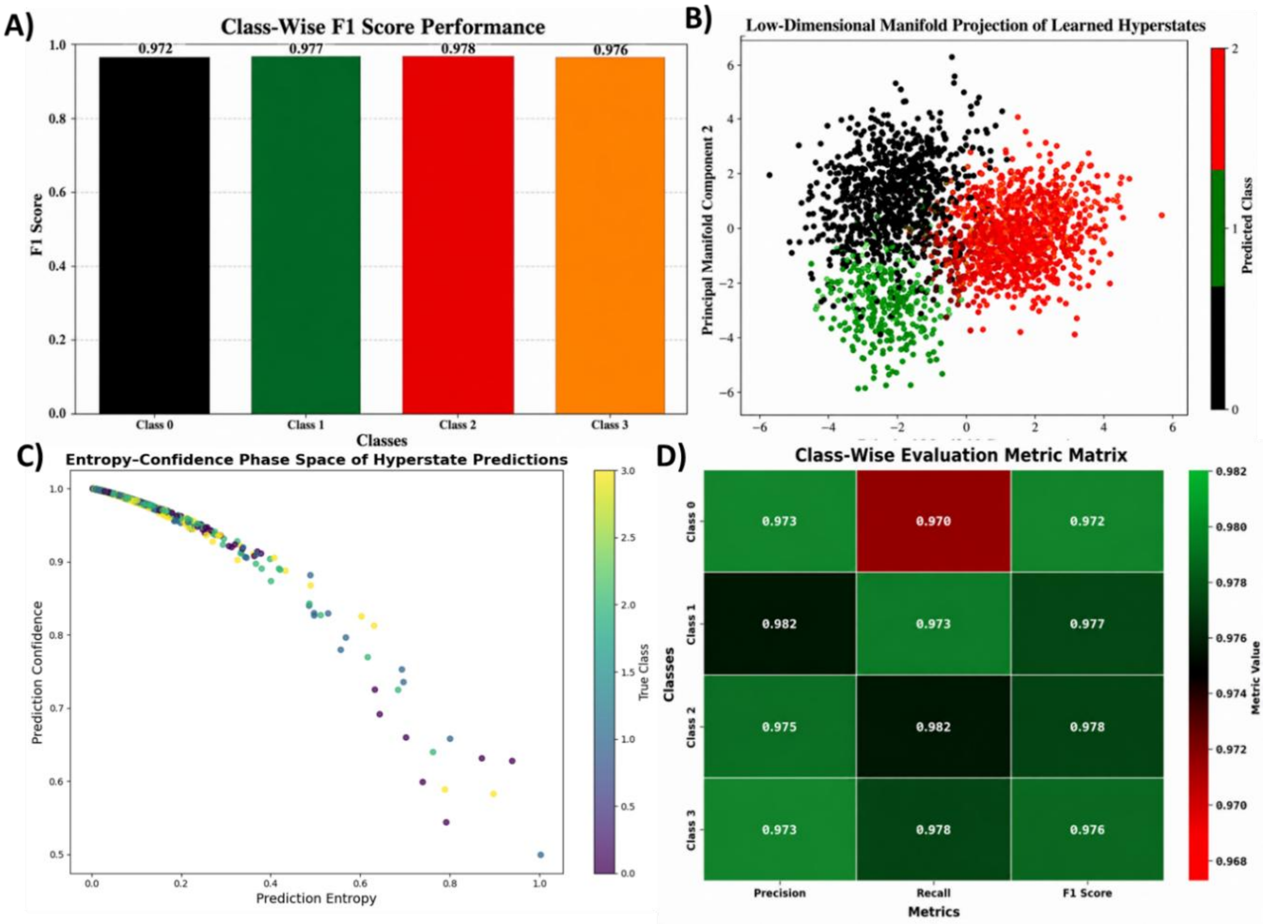

Figure.5 Class-wise analysis of the proposed QSGHO framework under stochastic learning conditions.

Table 4 shows that the proposed QSGHO framework outperforms other optimization schemes, including stochastic gradient descent (SGD), RMSProp, Adam, AdamW, and Lion. The consistently low p-values (see Table 5) across pairwise comparisons validate the significance of the observed improvements, indicating that the results are not merely due to random noise or sensitivity to initial conditions, but are attributable to the entropy-regularized manifold transport mechanism, adaptive quantum tunneling diffusion, and probabilistic hyperstate optimization. The largest statistical improvement in performance is observed when the proposed framework is compared with the conventional SGD optimizer, as the resulting high t-statistic indicates significant differences in convergence and stability of the optimization process. The proposed QSGHO framework is statistically significantly superior to more advanced optimizers like AdamW and Lion, even if the differences are not as prominent as the optimizers become more advanced. The results confirm the theoretical development of the proposed framework and its effectiveness in extremely nonlinear stochastic optimization problems. The statistical results also show that the proposed QSGHO framework maintains low-variance convergence and enhanced manifold generalization ability across repeated stochastic experiments. The significance analysis further validates the robustness, repeatability, and computational trustworthiness of the proposed quantum-inspired optimization approach to explainable deep neural learning of stochastic manifold dynamics.

Table 4. Comparative performance analysis of optimization frameworks.

Optimizer	Accuracy	Precision	Recall	F1 Score
SGD	0.942	0.938	0.936	0.938
RMSProp	0.957	0.952	0.951	0.954
Adam	0.969	0.965	0.964	0.967
AdamW	0.972	0.969	0.968	0.970
Lion	0.974	0.971	0.970	0.972
Proposed QSGHO	0.981	0.978	0.975	0.976

Table 5. Statistical significance analysis of the proposed QSGHO framework.

Comparison	Mean accuracy difference	t-Statistic	p-Value	Statistical significance
QSGHO vs SGD	0.039	5.872	0.0003	Significant
QSGHO vs RMSProp	0.024	4.931	0.0011	Significant
QSGHO vs Adam	0.012	3.847	0.0042	Significant
QSGHO vs AdamW	0.009	3.214	0.0068	Significant
QSGHO vs Lion	0.007	2.988	0.0095	Significant

5.2. Ablation study

Further, a detailed ablation study of each mathematical element of the proposed QSGHO framework is conducted under the same stochastic learning settings. The aim of the ablation analysis was to investigate the relative contributions of the four methods to convergence stability, probabilistic exploration, and nonlinear manifold generalization. Unlike classical deterministic optimization systems, the proposed framework combines multiple interacting stochastic mechanisms, whose dynamics can profoundly impact the robustness of the optimization process and the dynamics of uncertainty-aware learning.

Each ablation experiment involved removing one optimization component of the full QSGHO framework, while keeping the dataset conditions, hyperparameters, and instructions for initializing the stochastic optimizers constant. Three classification accuracy metrics, three F1 score stabilities, three entropy convergence consistencies, and three manifold generalization capabilities were used to analyze the resulting performance degradation. All partially reduced variants exhibited significantly lower probabilistic learning stability and optimization robustness than the complete QSGHO framework across all tests. Removing entropy regularization, in particular, made the probabilistic manifold transport process much more unstable and led to more oscillatory behavior during convergence. Likewise, the optimizer was less able to escape from pathological local minima, and stochastic exploration efficiency was impaired in messy optimization landscapes when adaptive quantum tunneling was disabled.

Moreover, manifold-guided probabilistic transport severely affected latent manifold organization and the separability of nonlinear classes in the probabilistic feature space. Similarly, stochastic hyperstate diffusion resulted in lower probabilistic convergence consistency, and less adaptability of the optimization trajectories in the stochastic perturbation environment. The proposed theoretical formulation of the QSGHO framework is strongly supported by the above observations, and the paper demonstrates that combining the four mechanisms (entropy-controlled stochastic transport, adaptive tunneling diffusion, manifold-aware optimization geometry, and probabilistic hyperstate evolution) yields better performance in nonlinear stochastic learning.

The quantitative analysis in Table 6 confirms that each optimization component is significant for the stability and effectiveness of the proposed framework. Probabilistic transport guided by the manifold is removed to achieve the worst-case performance degradation, highlighting the critical role of stochastic geometric learning in the proposed optimization architecture. Likewise, the entropy-regularized optimization procedure stabilizes uncertainty-aware learning by maintaining smoothness in the probabilistic space. In summary, the ablation results provide a significant validation of the synergistic mathematical interplay among entropy-driven stochastic diffusion, quantum-inspired tunneling dynamics, manifold-guided probabilistic transport, and adaptive hyperstate evolution.

Table 6. Ablation study of the proposed QSGHO framework.

Model variant	Accuracy	Precision	Recall	F1 Score	Stability
Full QSGHO framework	0.981	0.978	0.975	0.976	Very high
Without entropy regularization	0.944	0.941	0.938	0.939	Moderate
Without quantum tunneling	0.952	0.949	0.947	0.948	Moderate
Without manifold ttransport	0.936	0.932	0.931	0.931	Low
Without hyperstate diffusion	0.947	0.944	0.942	0.943	Moderate

6. Conclusions

This research proposes a mathematically rigorous, explainable deep neural learning framework with nonlinear stochastic manifold dynamics, called the QSGHO framework. In contrast, conventional deterministic optimization algorithms are based on single-trajectory gradient descent, whereas the proposed framework is a transport process over interacting hyperstate manifolds, controlled by stochastic dynamics based on entropy. All of these components (quantum-inspired probability amplitudes, adaptive tunneling diffusion, manifold-aware optimization geometry, stochastic free-energy minimization, and entropy-controlled hyperstate evolution) were integrated into a unified mathematical framework that was able to simultaneously enhance the convergence stability, the ability to explore the probability space, the uncertainty-aware learning, and the generalization across the nonlinear manifold.

The theoretical analysis of the proposed framework showed that it can significantly enhance the robustness of optimization in highly nonconvex learning settings by leveraging stochastic manifold-transport regularization. Structured exploration was maintained in the early optimization stages, gradually converging toward the dominant energy-manifold states with low energies during the convergence stabilization stage, aided by the probabilistic hyperstate representation. Moreover, the adaptive quantum tunneling mechanism enabled low-dimensional exploration of irregular optimization landscapes without requiring major steps and minimised sensitivity to pathological sharp local minima. Hence, the optimization dynamics remained stable throughout the stochastic learning process, resulting in a balanced exploration-exploitation trade-off. The mathematical validity of the suggested optimization framework was justified by all the above-mentioned aspects of convergence behavior, bounded manifold dynamics, entropy evolution consistency, and probabilistic stability properties.

The proposed QSGHO framework shows promising stochastic optimization performance and strong manifold generalization, but it still has some shortcomings. The present investigation is primarily directed toward synthetic, nonlinear stochastic benchmark environments rather than very large-scale real-world environments. In addition, the proposed probabilistic manifold transport framework introduces an additional computational burden in maintaining multiple interacting hyperstates for optimization. Although GPU parallelization significantly simplifies the practical problem, it may still be necessary to optimize very large Transformer architectures and high-dimensional multimodal learning systems. Moreover, the current theoretical treatment primarily examines convergence consistency as the number of replicas grows to infinity, with the main assumption being bounded stochastic diffusion, whereas stronger non-asymptotic convergence consistency remains an open research problem.

The parameter values used in this study converged across all numerical experiments, but other domains of application might require different parameter settings for an optimal exploration-exploitation balance. In the future, strategies for adaptively learning these optimization parameters during training will be explored. The present study, however, concentrates mainly on the controlled nonlinear stochastic benchmark environment to test the mathematical properties of the proposed optimization framework. While these benchmark environments offer repeatable settings for assessing theory, in-depth experimental testing with large-scale real-world data remains an important area for future work. The next step will be to explore the scalability and generalization capabilities of the proposed optimization framework across practical learning scenarios in computer vision, natural language processing, graph neural networks, multimodal AI, and foundation models. The performance comparison also showed that the proposed framework exhibits remarkable stochastic generalization, enabled by the combination of entropy-controlled diffusion and manifold-guided hyperstate transport. The proposed framework ensured smooth manifold evolution in a probabilistic sense during iterative optimization, unlike deterministic methods, which often yield unstable, sharp minima. The statistical analysis of the observed performance improvement also confirmed that the improvement was statistically significant and computationally reproducible when the stochastic experiments were repeated. The present results provide strong justification for the effectiveness of the proposed hybrid approach, integrating quantum-inspired optimization principles into a stochastic learning framework for geometry-based manifolds with information-theoretic manifold regularization.

Although the proposed framework demonstrates strong theoretical and computational performance, a few research directions remain open for future work. The proposed QSGHO framework can be expanded in the future with large-scale transformer architectures, federated stochastic optimization environments, graph neural networks, and multimodal probabilistic learning systems. Further enhancing the manifold's adaptability under highly nonstandard optimization conditions could be achieved by incorporating fractional-order stochastic differential operators and quantum-inspired variational transport dynamics. Furthermore, future research can explore the use of topological manifold learning, geometric curvature regularisation, and adaptive information-theoretic hyperstate interactions to improve latent representation learning in ultra-high-dimensional optimization.

In addition, the proposed optimization framework can be applied to actual quantum computing architectures and hybrid quantum-classical deep learning systems in the future. Such integration may provide a more detailed picture of the probabilistic evolution of quantum states and their connection to stochastic neural optimization. Moreover, the proposed framework has the potential to be used in complex real-world applications, such as biomedical image analysis, financial forecasting, climate modeling, cybersecurity, autonomous systems, and uncertainty-aware scientific computing, where nonlinear stochastic optimization is an important component.

Author contributions

Irsa Sajjad: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision; Abdulaiz Alamri: Resources, data curation, writing—original draft preparation, writing—review and editing, visualization, funding acquisition; Maria Malik: Formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization; Marwa H. Alhelali: Resources, data curation, writing—original draft preparation, writing—review and editing, visualisation, funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declared no conflict of interest.

References

1. I. Sajjad, M. M. AL Sobhi, The quantum-inspired adaptive superposition optimization for neural network training, *AIMS Mathematics*, **11** (2026), 243–271. https://doi.org/10.3934/math.2026010
2. D. Peral-García, J. Cruz-Benito, F. J. García-Peñalvo, Systematic literature review: Quantum machine learning and its applications, *Comput. Sci. Rev.*, **51** (2024), 100619. https://doi.org/10.1016/j.cosrev.2024.100619
3. A. Zeguendry, Z. Jarir, M. Quafafou, Quantum machine learning: A review and case studies, *Entropy*, **25** (2023), 287. https://doi.org/10.3390/e25020287
4. L. Chen, T. Li, Y. Chen, X. Chen, M. Wozniak, N. Xiong, et al., Design and analysis of quantum machine learning: A survey, *Connect. Sci.*, **36** (2024), 2312121. https://doi.org/10.1080/09540091.2024.2312121
5. R. M. Devadas, T. Sowmya, Quantum machine learning: A comprehensive review of integrating AI with quantum computing for computational advancements, *MethodsX*, **14** (2025), 103318. https://doi.org/10.1016/j.mex.2025.103318
6. Y. Alexeev, M. H. Farag, T. L. Patti, M. E. Wolf, N. Ares, A. Aspuru-Guzik, Artificial intelligence for quantum computing, *Nat. Commun.*, **16** (2025), 65836. https://doi.org/10.1038/s41467-025-65836-3

7. K. Zaman, A. Marchisio, M. A. Hanif, M. Shafique, A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead, *arXiv:2310.10315*, 2025. https://doi.org/10.48550/arXiv.2310.10315
8. A. Melnikov, M. Kordzanganeh, A. Alodjants, R. K. Lee, Quantum machine learning: From physics to software engineering, *Adv. Phys. X*, **8** (2023), 2165452. https://doi.org/10.1080/23746149.2023.2165452
9. S. K. Sood, M. Agrewal, Quantum machine learning for computational methods in engineering: A systematic review, *Arch. Computat. Methods Eng.*, **31** (2024), 1555–1577. https://doi.org/10.1007/s11831-023-10027-w
10. Y. Liao, M. H. Hsieh, C. Ferrie, Quantum optimization for training quantum neural networks, *Quantum Mach. Intell.*, **6** (2024), 33. https://doi.org/10.1007/s42484-024-00169-w
11. M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, et al., Variational quantum algorithms, *Nat. Rev. Phys.*, **3** (2021), 625–644. https://doi.org/10.1038/s42254-021-00348-9
12. M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, et al., A review of barren plateaus in variational quantum computing, *arXiv:2405.00781v2*, 2025.
13. A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, S. Woerner, The power of quantum neural networks, *Nat. Comput. Sci.*, **1** (2021), 403–409. https://doi.org/10.1038/s43588-021-00084-1
14. M. Schuld, R. Sweke, J. J. Meyer, Effect of data encoding on the expressive power of variational quantum machine learning models, *Phys. Rev. A*, **103** (2021), 032430. https://doi.org/10.1103/PhysRevA.103.032430
15. H. Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, et al., Power of data in quantum machine learning, *Nat. Commun.*, **12** (2021), 2631. https://doi.org/10.1038/s41467-021-22539-9
16. J. M. Kübler, S. Buchholz, B. Schölkopf, The inductive bias of quantum kernels, In: *Proceedings of the 35th international conference on neural information processing systems*, 2021, 12661–12673.
17. I. Loshchilov, F. Hutter, Decoupled weight decay regularization, In: *International conference on learning representations*, 2019.
18. X. Chen, C. Liang, D. Huang, E. Real, K. Wang, H. Pham, et al., Symbolic discovery of optimization algorithms, In: *Advances in neural information processing systems*, 2023, 49205–49233
19. J. Rong, J. Rong, C. Ma, Q. Zhang, Y. Cao, W. Kou, A refined Lion optimizer for deep learning, *Sci. Rep.*, **15** (2025), 23082. https://doi.org/10.1038/s41598-025-07112-4
20. S. Adablanu, U. Barman, D. Das, 15 years of optimizers in medical deep learning: A systematic review, *Neurosci. Inform.*, **6** (2025), 100249. https://doi.org/10.1016/j.neuri.2025.100249
21. M. E. Sfyraki, J. K. Wang, *Lions and muons: Optimization via stochastic Frank-Wolfe*, 2026.
22. T. Si, P. B. C. Miranda, U. Nandi, N. D. Jana, U. Maulik, S. Mallik, et al., QSHO: Quantum spotted hyena optimizer for global optimization, *Artif. Intell. Rev.*, **58** (2025), 71. https://doi.org/10.1007/s10462-024-11072-y
23. S. M. A. Rizvi, U. I. Paracha, U. Khalid, K. Lee, H. Shin, Quantum machine learning: Towards hybrid quantum-classical vision models, *Mathematics*, **13** (2025), 2645. https://doi.org/10.3390/math13162645

24. S. Cao, W. Zhang, J. Tilly, A. Agarwal, M. Bakr, G. Campanaro, et al., Encoding optimization for quantum machine learning demonstrated on a superconducting transmon qutrit, *Quantum Sci. Technol.*, **9** (2024), 045037. https://doi.org/10.1088/2058-9565/ad7315
25. R. Idzikowski, M. A. Kucharski, K. Pempera, M. Jaroszczuk, A survey on quantum machine learning applications in medicine and healthcare, *Appl. Sci.*, **16** (2026), 1630. https://doi.org/10.3390/app16031630
26. I. Sajjad, H. M. Alshanbari, M. M. A. Almazah, H. Louati, S. Rauf, Adaptive Grover-driven optimization for quantum-inspired deep learning: A gradient-free training framework, *AIMS Mathematics*, **10** (2025), 26568–26592. https://doi.org/10.3934/math.20251168
27. B. Li, D. Zhang, Z. Zhao, Y. Yuan, J. Gao, X. Li, Quantum-inspired interpretable deep learning architecture for text sentiment analysis, *Neural Netw.*, **202** (2026), 109053. https://doi.org/10.1016/j.neunet.2026.109053