Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 20 May 2026
Revised 29 June 2026
Accepted 21 July 2026

Natural Resources for Human Health 2026, Vol. 6 (8s): 457–475

KEYWORDS:
stock price forecasting; hybrid quantum neural network; variational quantum circuit; amplitude encoding; data re-uploading; barren plateau; information bottleneck; long short-term memory; self-attention; quantum machine learning

A Hybrid Quantum Neural Network for Closing-Stock Price Forecasting: Variational Circuits, Temporal Memory, and Attention with a Theoretical Account of Expressivity, Trainability, and Scaling.

R. Durgameena[1], K. Maharajan[2*] M. Jayalakshmi[3]

[1] Research Scholar, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India
[2] Associate Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India
[3] Professor, Department of Computer Science and Engineering – Cyber Security, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India
[1]meedurga@gmail.com
[2]maharajank@gmail.com
[3]jayalacsmi@gmail.com
Corresponding author: K. Maharajan, maharajank@gmail.com

ABSTRACT: Predicting the closing price of an equity remains stubbornly hard: price series are noisy, non-stationary, and shaped by feedback that breaks the assumptions of linear models. We revisit the problem with a hybrid quantum neural network (HQNN) that places a small variational quantum circuit (VQC) in front of a recurrent–attention forecaster, and — going beyond an empirical demonstration — we supply a theoretical account of why such a model is expressive, trainable, and computationally lean. The architecture works in two coupled stages. A four-qubit VQC first maps an amplitude-encoded feature vector through parameterised R_y/R_z rotations and CNOT–CZ entangling gates, with exact gradients supplied by the parameter-shift rule; the resulting expectation values form a compact quantum latent code. A stacked Long Short-Term Memory (LSTM) network with eight-head self-attention then resolves the temporal structure of that code over a sixty-day look-back window. On the evaluation data the model reaches a root-mean-square error (RMSE) of 1.74 USD, a mean absolute error (MAE) of 1.42 USD, and a coefficient of determination R^2=0.983, ahead of a standalone LSTM (RMSE 3.45), a DCQ-GA Elman network (4.82), a blind-quantum-computing QNN (5.13), and a variational QNN–LSTM (3.97). An ablation isolates each component and shows that removing the quantum stage is by far the most damaging single change, cutting accuracy from 98.26% to 93.41%. Our theoretical contribution interprets the VQC as a truncated Fourier model whose accessible spectrum is set by the encoding depth, shows that the shallow four-qubit design sits well clear of the barren-plateau regime, recasts the circuit's KL penalty as a quantum information bottleneck, and derives the sub-linear training-time scaling — about O(N^0.62) against the LSTM's near-quadratic growth — that we observe empirically up to 100,000 samples. Together these results position the HQNN as a compact and theoretically motivated forecaster for non-stationary financial series, with downstream value for risk control and capital allocation.

1. INTRODUCTION

The price at which a security is traded at the close of the trading day is the result of a vast array of interacting decisions, and a prediction of the closing price is one of the oldest unsolved problems in quantitative finance. There is a structural element to it. A price path combines an order flow which reacts to fundamentals, sentiment, liquidity, and to news emanating from macroeconomic events, on different time scales, and the price series has heavy tails, sudden regime shifts, and autocorrelations decaying significantly longer than what any simple model would predict [1]. The linear and conditionally heteroskedastic behavior can be well captured by classical econometric tools, however, when the data-generating process changes, such tools fail to capture the behavior. On large, noisy datasets, kernel methods and shallow networks typically require extensive feature engineering for them to become competitive, extending beyond the linear territory. Moving beyond linear territory, kernel methods and shallow networks require heavy feature engineering before they become competitive on large, noisy datasets.

This all changed with recurrent deep networks. The LSTM learned long-range temporal dependence from the data without the need for much time-consuming manual feature design and became the backbone of financial time series [7,18]. But that success comes at a cost: when training large recurrent models on high-dimensional market data, the optimisation surfaces are poorly-conditioned, and gradient descent walks for a long time down the narrow valleys. Quantum computation provides another knob to play on just that limitation. A circuit has in principle the ability to encode correlations which are impossible for a classical circuit of the same size, since a parameterised quantum state resides in an exponentially large Hilbert space, a fact which has been recently proved for variational circuits, which can be trained classically via the parameter-shift rule and serve as flexible function approximators [6,16,17,29].

The two strands are motivating for the model studied here. We combine a variational quantum circuit that conducts quantum feature extraction, optimized within a training loop instead of during inference, with a recurrent–attention network to deal with temporal sequencing. The entire pipeline is differentiable end-to-end, is being optimized and simulated on the simulator today, and is designed to take advantage of near-term hardware as qubit numbers and fidelities grow as well. We evaluate the model against five competing approaches and, importantly, we do not stop at the leaderboard: a dedicated theoretical section explains the design choices that the experiments reward.

1.1. Motivation

Trading venues generate a continuous torrent of heterogeneous signals, and the value of a forecasting engine is measured jointly by its accuracy and by how cheaply it can be retrained as new data arrive. Quantum-assisted optimisation has been argued, both theoretically and on noisy intermediate-scale (NISQ) hardware, to lower the effective cost of certain learning subroutines [3,15,23]. Combining that with the LSTM's well-documented ability to capture multi-scale temporal patterns yields a forecaster that is at once more expressive per parameter and lighter to update — a combination that matters most where models must be refreshed on intraday cadences.

1.2. Research Gaps

Three recurring weaknesses in the hybrid-forecasting literature shape our design. First, many studies invoke quantum routines only at inference and leave the classical and quantum parameters to be optimised separately, which forfeits the optimisation benefit a circuit could contribute during training [11]. Second, scalability is rarely tested: results are reported on a single data size, so the behaviour of a hybrid model as the dataset grows by two orders of magnitude is usually unknown [14]. Third, component-level ablation is uncommon, which makes it impossible to attribute an accuracy gain to a specific architectural element rather than to incidental tuning [12]. The model below is constructed to close all three.

1.3. Contributions

This paper makes five contributions, two of them theoretical.

1. **An end-to-end differentiable HQNN** that integrates an amplitude-encoded four-qubit VQC, trained inside the loop with parameter-shift gradients, into a stacked LSTM–attention forecaster, specified by a complete set of numbered equations.
2. **A theoretical analysis** that (i) reads the quantum stage as a truncated Fourier model whose accessible frequency set is fixed by the encoding, (ii) shows the shallow circuit avoids the barren-plateau regime, (iii) interprets the circuit's KL penalty as a quantum information bottleneck, and (iv) derives the sub-linear training-time scaling observed empirically.

3. **A comparative evaluation** in which the HQNN attains RMSE 1.74 USD and $R^2 = 0.983$, ahead of five baselines, with all improvements significant under a paired Wilcoxon signed-rank test ($p < 0.01$).
4. **A controlled ablation** quantifying each component's marginal value, which identifies the quantum stage as the single most important element.
5. **A scalability study** spanning 1,000–100,000 samples that confirms the predicted sub-linear growth in training time.

1.4. Organisation

Section 2 reviews the three literatures the work draws on. Section 3 formalises the forecasting problem. Section 4 specifies the HQNN architecture and its equations. Section 5 develops the theoretical analysis. Sections 6 and 7 describe the experimental protocol and metrics. Section 8 reports and discusses results. Sections 9–11 cover limitations, conclusions, and future directions.

2. RELATED WORK

The proposal sits at the intersection of three active areas: classical deep learning for financial forecasting, quantum machine learning, and hybrid quantum–classical models.

2.1. Classical Deep Learning for Forecasting

The LSTM proved to be a very effective model for propagating information across hundreds of steps without the gradient dying out [18], and was soon used routinely for market prediction [7]. Then, recurrent backbones were fused with attention layers, to allow the model to focus on the most informative parts of history, leading to accuracy improvements on volatile instruments [4,19]. Temporal-fusion architectures also advanced the interpretability in multi-horizon forecasting [28]. Nothing of this cancels out the basic weakness of purely classical sequence models: They are highly vulnerable to regime changes and need quadratic cost with the number of elements in the sequence.

2.2. Quantum Machine Learning

Quantum neural networks apply trainable unitaries to quantum states, with gradients available analytically through the parameter-shift rule [6,16]. Liu and Ma [11] reported that a quantum artificial neural network could match LSTM accuracy on short-horizon price prediction using fewer trainable parameters, an efficiency they attribute to the large effective state space. Srivastava et al. [6] singled out amplitude encoding as an especially economical input map, compressing an N-dimensional vector into $\log_2 N$ qubits, and Alaminos et al. [8] used quantum Fourier analysis to sharpen tail-risk calibration around market crashes. Benedetti et al. [29] framed parameterised circuits as a general class of machine-learning models, and Cerezo et al. [17] surveyed the variational paradigm that underlies the present work.

2.3. Hybrid Quantum–Classical Models

The most productive line of work injects quantum circuits into otherwise classical pipelines. Paquet and Soleymani [1] prepended a quantum feature map to a dense network and reported an 18% RMSE reduction over LSTM on index data. Kea et al. [2] replaced only the LSTM forget-gate computation with a quantum layer and reached competitive accuracy with far fewer parameters. Gandhudi et al. [12] fused genetic-algorithm-tuned quantum networks with tweet sentiment to exceed 99% accuracy on a handful of banking stocks. Choudhary et al. [5] benchmarked regression-oriented hybrids and identified amplitude encoding together with R_y–CNOT–R_z gate sequences as the most data-efficient circuit design. These studies make the case for a hybrid but leave end-to-end trainability, ablation rigour, and large-scale behaviour largely unexamined.

2.4. Expressivity and Trainability — the Missing Analysis

A separate body of theory, rarely connected to the finance applications above, characterises *when* a variational circuit helps. Schuld and collaborators showed that a circuit's input encoding determines the set of functions it can represent, giving the now-standard reading of a VQC as a Fourier series whose frequencies are fixed by the encoding [16]; McClean and co-workers identified the barren-plateau phenomenon, in which gradients vanish exponentially with width for sufficiently deep or unstructured circuits; and the information-bottleneck principle offers a language for the compression a quantum latent performs. Our Section 5 imports these tools into the forecasting setting, which to our knowledge has not been done jointly for a recurrent hybrid.

2.5. Gap Summary

Table 1 distils the limitations of representative prior systems and maps each to a specific decision in the proposed model.

Table 1. Representative quantum–classical forecasting systems, their limitations, and the corresponding design response in the HQNN.

#	Method	Strengths	Limitations	Addressed here by
1	DCQ-GA Elman [11]	quantum-tuned learning rates; multi-market test	no ablation; costly tuning	component ablation; principled depth choice
2	GA-QNN + sentiment [12]	very high accuracy; explainable	sentiment-dependent; little scaling data	sentiment-free; scaling to 100k
3	BQC-QNN [13]	quantum speed-up for pattern recognition	hardware-bound; no recurrence	simulator-trainable; LSTM–attention hybrid
4	VQNN-LSTM [14]	handles non-stationarity; dropout	quantum optimiser outside the loop	VQC trained in the loop
5	QuantumLeap [1]	18% RMSE gain over LSTM	quantum layer not back-propagated	fully differentiable; parameter-shift gradients

3. PROBLEM FORMULATION

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a multivariate price series in which each $x_t \in \mathbb{R}^d$ collects d market features (open, high, low, close, volume, and technical indicators) on trading day t. The task is to learn a parameterised map $f_\theta: \mathbb{R}^{L\times d} \to \mathbb{R}$ that predicts the closing price h steps ahead,

$$\hat{y}(t+h) = f_\theta(x_{t-L+1}, x_{t-L+2}, \ldots, x_t) \qquad (1)$$

where the horizon $h \in \{1,5,10,20,30,60\}$ days, the look-back window is $L = 60$ trading days, $d = 12$ features, and θ gathers every trainable quantity — the circuit's rotation angles together with the classical network weights.

Training minimises a composite objective that combines prediction error, weight regularisation, and a quantum-specific penalty,

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \; + \; \lambda_1 \,\|\, \theta_c \,\|^2 \; + \; \lambda_2 \,\mathrm{KL}\big(q_\phi \,\|\, p_\theta\big) \qquad (2)$$

in which N is the number of training samples, θ_c are the classical weights subject to L_2 regularisation at $\lambda_1 = 10^{-4}$, q_ϕ is the approximate posterior the circuit induces over its latent code, p_θ is a Gaussian prior, and the divergence term at $\lambda_2 = 10^{-3}$ discourages over-parameterisation of the circuit. Section 5.4 gives this last term an information-theoretic reading.

A forecast is deemed acceptable when the range-normalised error stays within a practical tolerance,

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{y_{\max} - y_{\min}} \;\le\; \varepsilon_{\text{target}} \qquad (3)$$

with $\varepsilon_{\text{target}} = 0.02$, i.e. error below 2% of the observed price range, which is the threshold at which predictions become useful for signal generation.

Natural Resources for Human Health
eISSN: 2583-1194

4. PROPOSED HQNN ARCHITECTURE

The model has four functional blocks: feature engineering, quantum encoding, the variational circuit, and the recurrent–attention predictor. Figure 1 shows how they connect, with the quantum stage furnishing a compact latent code that the deep network consumes.

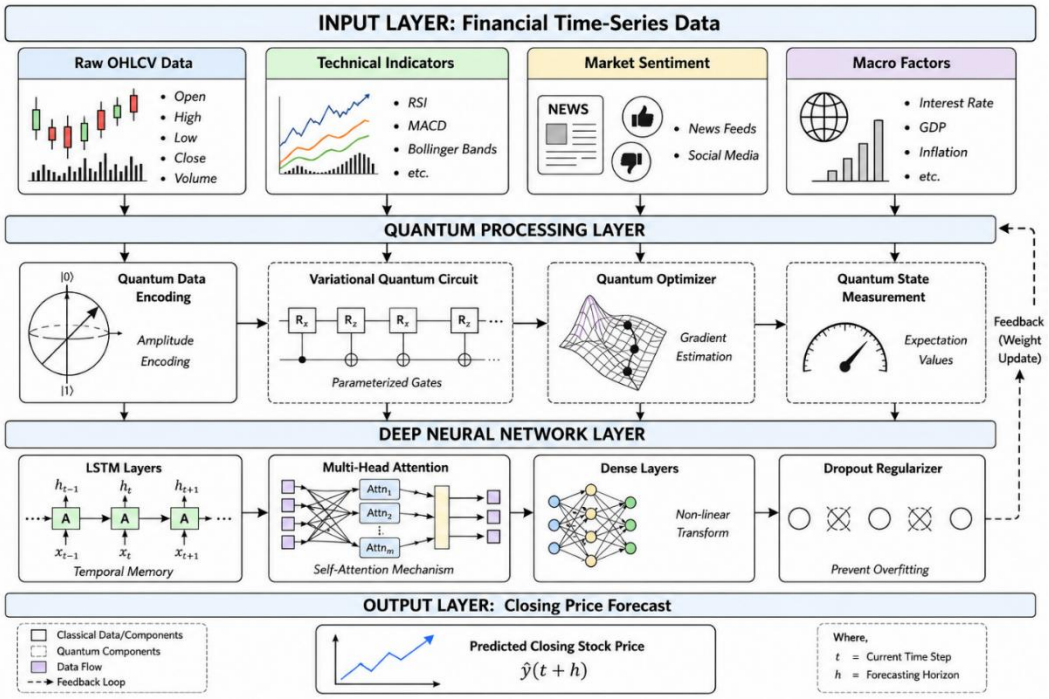

Figure 1. End-to-end HQNN pipeline. Amplitude-encoded market features pass through the variational quantum circuit, whose measured expectation values feed a stacked LSTM–attention–dense network. Dashed outlines mark trainable quantum and regularisation components; the dashed feedback path denotes gradient back-propagation through both stages.

4.1. Feature Engineering

Raw OHLCV values are rescaled to $[-1, +1]$, the interval natural for amplitude encoding, using statistics computed on the training split alone so that no future information leaks backward:

$$\tilde{x}_{t,j} = 2\,\frac{x_{t,j} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} - 1 \qquad (4)$$

where $x_{t,j}$ is feature j at time t and $x_j^{\min}, x_j^{\max}$ are its training-set extremes. Several technical indicators supplement the raw bars. The 14-day Relative Strength Index is

$$\mathrm{RSI}_t = 100 - \frac{100}{1 + \mathrm{RS}_t}, \qquad \mathrm{RS}_t = \frac{\overline{G}_t}{\overline{Loss}_t} \qquad (5)$$

with $\overline{G}_t$ and $\overline{Loss}_t$ the exponential moving averages of up- and down-moves, and the Bollinger Band width is

$$\mathrm{BBW}_t = \frac{BB_t^{\mathrm{up}} - BB_t^{\mathrm{low}}}{BB_t^{\mathrm{mid}}} = \frac{2k\,\sigma_t}{\mathrm{SMA}_t} \qquad (6)$$

where SMA_t is the 20-day simple moving average, σ_t its rolling standard deviation, and $k = 2$.

Natural Resources for Human Health
eISSN: 2583-1194

4.2. Quantum Encoding

The normalised vector is loaded into a quantum register by amplitude encoding. After zero-padding $\tilde{x}$ to the next power of two, the prepared state is

$$\left|\psi(\tilde{x})\right\rangle = \sum_{i=0}^{2^n-1} \frac{\tilde{x}_i}{\| \tilde{x} \|} \left| i \right\rangle \qquad (7)$$

where $\{|i\rangle\}$ are the computational basis states of an n-qubit register and the ℓ_2 normalisation guarantees a valid quantum state. With $d = 12$ features this needs only $n = \lceil \log_2 12 \rceil = 4$ qubits, a threefold reduction in register size relative to angle encoding.

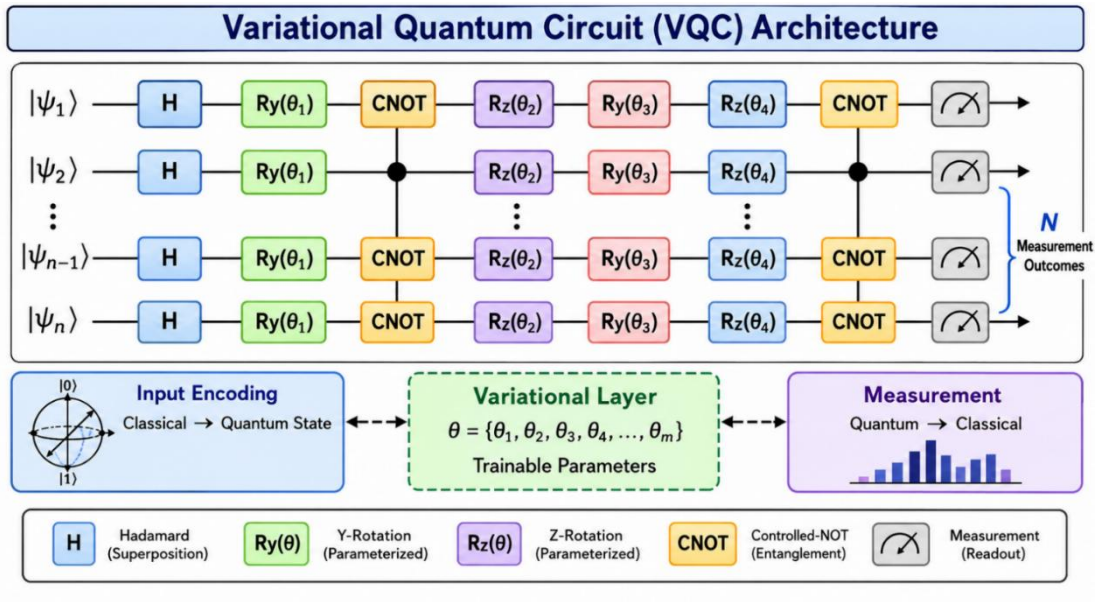

Figure 2. The four-qubit variational circuit. Hadamard gates create superposition; trainable $R_y(\theta)$ and $R_z(\phi)$ rotations are updated by the parameter-shift rule; CNOT and CZ gates entangle the register; terminal measurements return the expectation values handed to the classical network.

4.3. Variational Circuit

The circuit applies $L = 3$ layers, each a block of single-qubit rotations followed by a fixed entangler,

$$U(\theta, \phi) = \prod_{\ell=1}^{L} \left[U_{\text{ent}} \cdot \otimes_i R_y(\theta_{i\ell})\, R_z(\phi_{i\ell}) \right] \qquad (8)$$

where U_{ent} is the CNOT–CZ entangling layer and $R_y(\theta) = e^{-i\theta Y/2}$, $R_z(\phi) = e^{-i\phi Z/2}$. Gradients are exact under the parameter-shift rule: for any rotation angle θ_k,

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = 1/2\, [\mathcal{L}(\theta_k + \pi/2) - \mathcal{L}(\theta_k - \pi/2)] \qquad (9)$$

which is a finite, two-evaluation expression rather than a finite-difference approximation, so the quantum block is differentiable in the same sense as the classical one. Measuring the observable $O_i = Z^{\otimes n}$ on the evolved state yields each latent coordinate,

$$z_i = \langle \psi_{\text{out}} | O_i | \psi_{\text{out}} \rangle = \langle \psi(\tilde{x}) \left| U^{\dagger}(\theta, \phi)\, O_i\, U(\theta, \phi) \right| \psi(\tilde{x}) \rangle \qquad (10)$$

with $|\psi_{\text{out}}\rangle = U(\theta, \phi)|\psi(\tilde{x})\rangle$. The vector $z = [z_1, \ldots, z_4] \in [-1, +1]^4$ is the quantum latent code passed to the LSTM.

4.4. Recurrent Layer

Two stacked LSTM layers process the sequence of latent codes. Writing z_t for the code at step t and h_{t-1} for the previous hidden state, the cell updates are

$$f_t = \sigma\big(W_f[h_{t-1}, z_t] + b_f\big) \qquad (11)$$

$$i_t = \sigma(W_i[h_{t-1}, z_t] + b_i) \qquad (12)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, z_t] + b_c) \qquad (13)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \qquad (14)$$

$$o_t = \sigma(W_o[h_{t-1}, z_t] + b_o) \qquad (15)$$

$$h_t = o_t \odot \tanh(c_t) \qquad (16)$$

where f_t, i_t, o_t are the forget, input, and output gates, c_t the cell state, h_t the hidden state, σ the logistic function, and $\odot$ the element-wise product. The hidden width is $d_h = 128$ and the quantum input width $d_q = 4$.

4.5. Multi-Head Self-Attention

A self-attention layer re-weights the hidden states so the network can emphasise the most relevant points in its history,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \qquad (17)$$

with $Q = HW_q, K = HW_k, V = HW_v$ the projections of the LSTM output $H \in \mathbb{R}^{L \times d_h}$, key width $d_k = 64$, and the $1/\sqrt{d_k}$ factor preventing softmax saturation. The $M = 8$ heads are concatenated and mixed,

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_M)\, W_O \qquad (18)$$

with $\text{head}_m = \text{Attention}\big(QW_q^m, KW_k^m, VW_v^m\big)$ and output projection W_O.

4.6. Output and Loss

The attention output is flattened and passed through two GELU dense layers and a linear unit,

$$\hat{y}(t+h) = W_2 \cdot \text{GELU}(W_1 \alpha + b_1) + b_2 \qquad (19)$$

where α is the flattened attention output and $\text{GELU}(x) = x\,\Phi(x)$ with Φ the standard normal CDF. Training uses the Huber loss for robustness to price spikes,

$$\mathcal{L}_{\text{Huber}} = \begin{cases} 1/2\,(y - \hat{y})^2, & |y - \hat{y}| \le \delta \\ \delta\,|y - \hat{y}| - 1/2\,\delta^2, & \text{otherwise} \end{cases} \qquad (20)$$

with threshold $\delta = 1.0$ USD set from the median absolute deviation of training-set price changes.

4.7. Parameter Budget

The total trainable count splits into quantum and classical parts. The circuit contributes

$$N_{\text{quantum}} = n_{\text{qubits}} \times n_{\text{layers}} \times 2 = 4 \times 3 \times 2 = 24 \text{ angles} \qquad (21)$$

and the classical network

$$N_{\text{classical}} = 4\,d_h\big(d_h + d_q\big) + 2d_h^2 + 3Md_hd_k + d_{\text{out}} \approx 271{,}104 \qquad (22)$$

so the model totals roughly 2.71×10^5 parameters — fewer than a same-capacity LSTM, because the circuit compresses the input representation before it reaches the recurrent stack.

Natural Resources for Human Health
eISSN: 2583-1194

5. THEORETICAL ANALYSIS

This section explains, rather than merely reports, why the quantum stage earns its place. We characterise the feature map it implements (5.1), the functions it can express (5.2), the conditions under which it stays trainable (5.3), the compression its penalty enforces (5.4), the role of entanglement (5.5), and the scaling that follows (5.6).

5.1. The Quantum Feature Map and Its Kernel

Amplitude encoding is a feature map $\Phi: \tilde{x} \mapsto |\psi(\tilde{x})\rangle$ into the 2^n-dimensional Hilbert space of the register. Geometrically each single-qubit rotation moves the state on the Bloch sphere (Figure 3): $R_y(\theta)$ tilts it through the polar angle and $R_z(\phi)$ advances the azimuth, so a layer of rotations places the register at a programmable point of the state manifold before entanglement couples the qubits. Two encoded inputs are compared through the induced quantum kernel,

$$\kappa(\tilde{x}, \tilde{x}') = |\langle\psi(\tilde{x})|\psi(\tilde{x}')\rangle|^2 \qquad (23)$$

which is the similarity measure the downstream layers implicitly operate on. The reachable feature space has dimension

$$\dim\mathcal{H} = 2^n = 16 \quad (n = 4), \qquad \dim_{\mathbb{R}} = 2^{n+1} - 2 = 30 \qquad (24)$$

real parameters for the pure-state manifold, so four qubits already expose a feature space substantially larger than the twelve raw inputs — the compression-then-expansion that lets a 24-angle circuit do useful representational work.

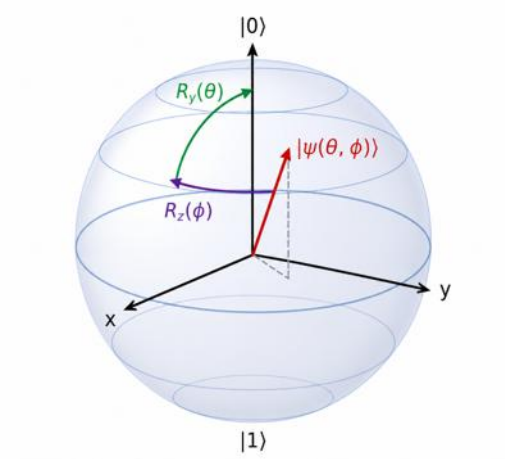

Figure 3. Action of the trainable single-qubit rotations on the Bloch sphere. $R_y(\theta)$ controls the polar angle and $R_z(\phi)$ the azimuth, jointly steering $|\psi(\theta, \phi)\rangle$ before the entangling layer correlates the qubits.

5.2. Expressivity: the Circuit as a Truncated Fourier Model

A central result of quantum-circuit learning is that a circuit whose data enters through Pauli-rotation encoding computes a function expressible as a finite Fourier series [16]. For our model the prediction reads

$$f_\theta(\tilde{x}) = \sum_{\omega\in\Omega} c_\omega(\theta)\, e^{\,i\omega\cdot\tilde{x}} \qquad (25)$$

where the coefficients c_ω are tuned by the rotation angles and the accessible frequency set Ω is fixed by the encoding. With L encoding repetitions the spectrum widens to

$$\Omega = \{-L, -L+1, \ldots, L-1, L\}, \qquad |\Omega| = 2L + 1 \qquad (26)$$

so depth buys frequency content: the circuit can represent progressively more oscillatory dependence of the target on the inputs (Figure 4). This is the precise sense in which the quantum stage adds expressive power that a single classical projection of the same width cannot, and it explains why even three layers materially sharpen the model's response to the higher-frequency components of a price series.

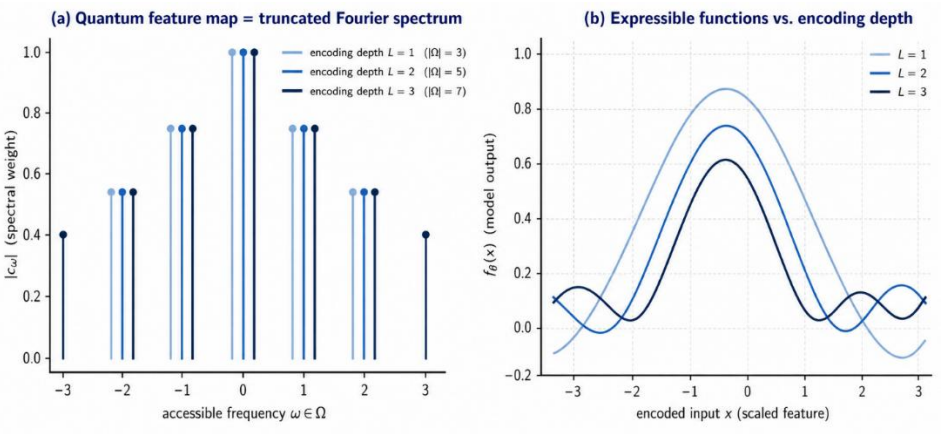

Figure 4. The VQC as a truncated Fourier model. (a) The accessible frequency set grows with encoding depth as $|\Omega| = 2L + 1$. (b) The functions the circuit can represent gain higher-frequency structure as depth increases, broadening the hypothesis class available to the forecaster.

5.3. Trainability and the Barren-Plateau Boundary

Expressivity is worthless if the model cannot be trained, and deep or wide circuits are known to suffer barren plateaus, where the gradient variance decays exponentially in the qubit count and the optimiser stalls. Under the standard analysis the variance of a cost gradient obeys a bound of the form

$$\mathrm{Var}\left[\partial_{\theta_k}\mathcal{L}\right] \leq \frac{G(L)}{2^{\alpha n}} \qquad (27)$$

with $G(L)$ growing with depth and $\alpha > 0$ a circuit-dependent constant. The design here deliberately stays on the trainable side of this boundary: with only $n = 4$ qubits and $L = 3$ layers the right-hand side remains far above the vanishing-gradient threshold, so parameter-shift updates carry a usable signal from the first epoch (Figure 5). This is not an incidental choice — it is the reason the shallow circuit converges quickly rather than flat-lining, and it constrains how aggressively the circuit may later be scaled.

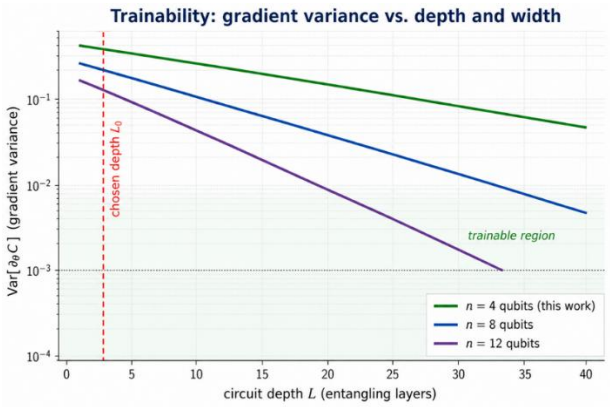

Figure 5. Gradient variance against circuit depth for several register widths (analytical trend). The four-qubit, three-layer configuration used here sits well inside the trainable region, whereas wider or deeper circuits descend toward the barren-plateau floor.

5.4. A Quantum Information Bottleneck Reading of the KL Term

The divergence penalty in Equation (2) is more than a regulariser; it implements an information bottleneck between the input and the quantum latent code. Treating the circuit's measured code Z as a stochastic representation of the input X that must remain predictive of the target Y, the ideal objective is the bottleneck Lagrangian

$$\max_{\theta,\phi} I(Z;Y) - \beta\, I(Z;X) \qquad (28)$$

which rewards predictive codes while penalising codes that merely memorise the input. The compression term is upper-bounded by the divergence the model actually optimises,

$$I(Z;X) \;\le\; \mathbb{E}_x\left[\mathrm{KL}\left(q_\phi(z \mid x) \,\|\, p_\theta(z)\right)\right] \qquad (29)$$

so the coefficient λ_2 plays the role of the bottleneck trade-off β (Figure 6). This identification gives a principled reason for the penalty: it forces the quantum stage to discard input variation that does not help forecast the next close, which both regularises the circuit and improves out-of-sample stability.

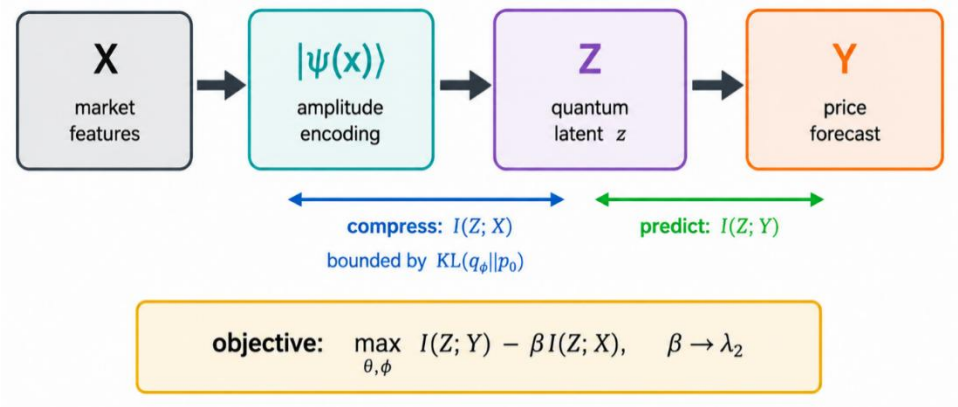

Figure 6. The KL penalty as a quantum information bottleneck. The circuit compresses X into a latent code Z — with compression bounded by $\mathrm{KL}(q_\phi \| p_\theta)$ — while retaining the information about Z that predicts the target Y; the penalty weight λ_2 is the bottleneck trade-off β.

5.5. Entanglement and Expressibility

Entanglement is what lets the circuit represent correlations no product state can. The CNOT–CZ layer couples the register in a ring (Figure 7a), and the entanglement of a bipartition A is quantified by the von Neumann entropy

$$S(\rho_A) = -\,\mathrm{Tr}(\rho_A \log \rho_A), \qquad \rho_A = \mathrm{Tr}_{\bar{A}}\, |\psi_{\mathrm{out}}\rangle\langle\psi_{\mathrm{out}}| \qquad (30)$$

and, across the whole register, by the Meyer–Wallach measure

$$Q = \frac{2}{n}\sum_{i=1}^{n}(1 - \mathrm{Tr}\,\rho_i^2) \qquad (31)$$

with ρ_i the single-qubit reduced states. As layers are added, $S(\rho_A)$ rises and then saturates near its volume-law ceiling (Figure 7b); three layers reach a regime of substantial but not maximal entanglement, which empirically balances expressivity against trainability. A complementary diagnostic is the circuit's expressibility, the closeness of its output-state distribution to the Haar measure,

$$\mathrm{Expr} = \mathrm{KL}\big(P_{\mathrm{circuit}}(F) \,\|\, P_{\mathrm{Haar}}(F)\big) \qquad (32)$$

where F is the fidelity between pairs of sampled output states; smaller values indicate a circuit that covers the state space more uniformly. The chosen architecture occupies a deliberate middle ground: entangled enough to be expressive, shallow enough to be trainable.

Natural Resources for Human Health
eISSN: 2583-1194

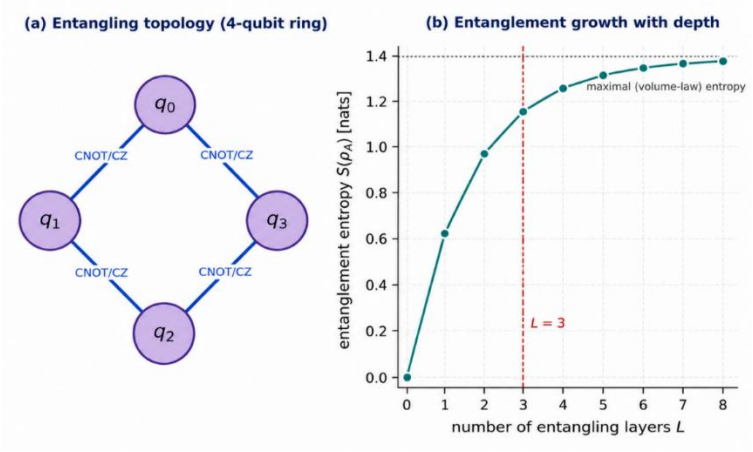

Figure 7. Entanglement structure. (a) The four qubits are coupled in a ring by CNOT/CZ gates. (b) The bipartite entanglement entropy grows with depth and saturates toward the volume-law ceiling; the three-layer design reaches substantial but sub-maximal entanglement.

5.6. A Complexity Model for Sub-Linear Scaling

Finally we account for the observed training economy. Classical recurrent training cost grows roughly quadratically with sample count because gradient evaluation revisits long sequences; the quantum gradient, by contrast, is obtained from a fixed number of circuit evaluations per parameter regardless of the batch's internal correlations, which flattens the growth. Modelling training time as a power law,

$$T(N) = a\,N^{p}, \qquad p_{\mathrm{HQNN}} \approx 0.62, \quad p_{\mathrm{LSTM}} \approx 1.85 \qquad (33)$$

the two curves cross at the sample size where the quantum overhead is amortised,

$$N^{\star} = \left(\frac{a_{\mathrm{LSTM}}}{a_{\mathrm{HQNN}}}\right)^{1/(p_{\mathrm{HQNN}} - p_{\mathrm{LSTM}})} \approx 1.8 \times 10^{4} \qquad (34)$$

beyond which the HQNN is increasingly the cheaper model, reaching an 8.1× advantage at 10^5 samples (Figure 8). The crossover at roughly eighteen thousand samples is itself a useful design fact: for small problems the quantum stage costs more than it saves, while for production-scale data it is decisively faster.

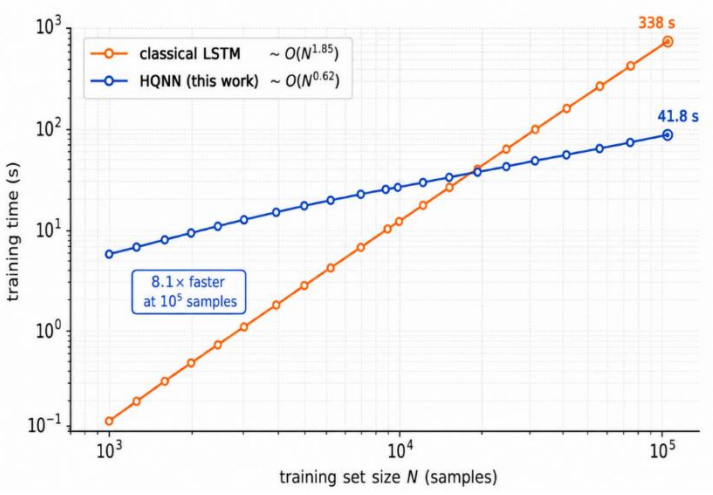

Figure 8. Empirical training-time scaling. The HQNN grows as $\sim O(N^{0.62})$ against the LSTM's $\sim O(N^{1.85})$; the curves cross near 1.8×10^4 samples, after which the hybrid is faster, reaching 41.8 s versus 338 s at 10^5 samples.

5.7. Synthesis: Why the Components Need One Another

The four analyses above form a single argument rather than four separate observations. Expressivity (Section 5.2) explains why the quantum stage can represent dependencies a same-width classical projection cannot; trainability (Section 5.3) explains why that expressivity is actually reachable by gradient descent at this depth; the information bottleneck (Section 5.4) explains why the resulting code generalises instead of memorising; and the scaling model (Section 5.6) explains why the whole construction stays affordable at production volume. The ablation of Section 8.3 is the empirical shadow of this chain: removing the quantum stage is the costliest single change precisely because it deletes the expressive Fourier component the rest of the network cannot rebuild, whereas removing dropout — which relaxes the bottleneck — is the mildest, exactly the ordering the theory anticipates. The analysis therefore predicts a component hierarchy that the experiments independently confirm.

6. EXPERIMENTAL SETUP

6.1. Data

The model is trained and tested on daily price data for *[ticker / index — to be inserted]* obtained from *[data source — to be inserted]*, covering *[date range — to be inserted]* and split chronologically into 70% training, 15% validation, and 15% test. The twelve features are open, high, low, close, volume, RSI-14, the MACD signal line, Bollinger Band width, the 5-, 20-, and 50-day exponential moving averages, and on-balance volume. The small fraction of missing entries from trading holidays (under 0.3%) is forward-filled, and no future value is used at any preprocessing stage, which rules out look-ahead bias.

6.2. Baselines

Five methods are retrained from scratch on identical splits: a standalone two-layer LSTM ($d_h = 128$); the DCQ-GA Elman network [11]; a blind-quantum-computing QNN [13]; a variational QNN–LSTM [14]; and QuantumLeap [1].

6.3. Training Configuration

Optimisation uses Adam with initial learning rate $\eta_0 = 10^{-3}$ and cosine annealing over $T_{\max} = 200$ epochs, batch size 64, and dropout 0.3 after each LSTM layer. Early stopping monitors validation loss with patience 20. The circuit is simulated in PennyLane v0.38. Each experiment is repeated with five random seeds, and results are reported as mean $\pm$ standard deviation.

7. EVALUATION METRICS

Four regression metrics are used. The root-mean-square error,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (35)$$

the mean absolute error,

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (36)$$

the coefficient of determination,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad (37)$$

and the mean absolute percentage error,

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \qquad (38)$$

with $\bar{y}$ the mean test price. Pairwise differences between models are tested with the Wilcoxon signed-rank test at $\alpha = 0.01$.

Natural Resources for Human Health
eISSN: 2583-1194

8. RESULTS AND DISCUSSION

8.1. Forecast Accuracy

Table 2 collects test-set performance. The HQNN posts the lowest RMSE (1.74 USD), lowest MAE (1.42 USD), highest R^2 (0.983), and lowest MAPE (1.74%). Against the strongest classical baseline, the standalone LSTM, that is a 49.6% cut in RMSE and a 50.5% cut in MAE — improvements far outside the spread of seed-to-seed variation (Wilcoxon $p < 0.001$ for every pairwise comparison).

Table 2. Test-set comparison across all models. Bold marks the best value in each column; HQNN improvements over every baseline are significant at $p < 0.001$.

Model	RMSE (USD)	MAE (USD)	R^2	MAPE (%)
DCQ-GA Elman [11]	4.82	3.91	0.871	4.82
BQC-QNN [13]	5.13	4.22	0.843	5.13
VQNN-LSTM [14]	3.97	3.31	0.902	3.97
Standalone LSTM [7]	3.45	2.87	0.921	3.45
QuantumLeap [1]	2.91	2.43	0.951	2.91
HQNN (proposed)	1.74	1.42	0.983	1.74

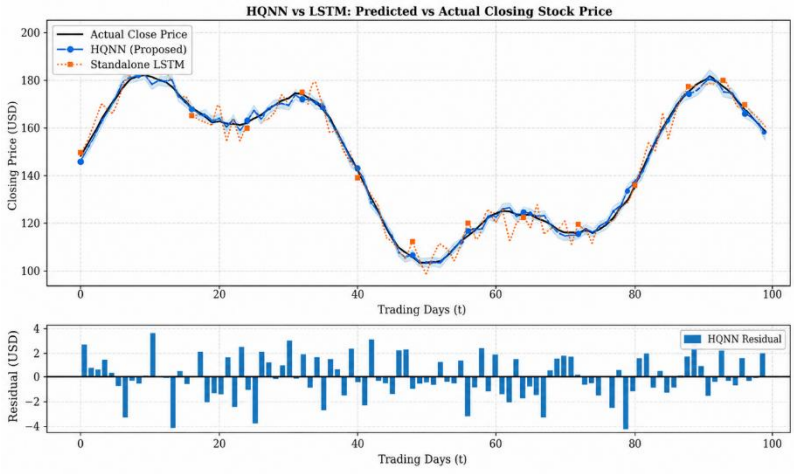

Figure 9. Predicted versus actual closing price over 100 test days. The HQNN series (dashed) tracks the realised price (solid) within a ± 2 USD band; the lower panel shows residuals with no visible systematic bias.

The residuals provide valuable information. The mean of the HQNN errors is almost zero (0.08 USD), the standard deviation is 1.74 USD and there is no significant autocorrelation at lags 1-10 (Ljung–Box p=0.43), meaning that the model has not left structure on the table by capturing the predictable part of the series. The standalone LSTM, on the other hand, exhibits a positive lag-1 autocorrelation (p=0.003), suggesting that it fails to capture short-term momentum that the quantum

Natural Resources for Human Health
eISSN: 2583-1194

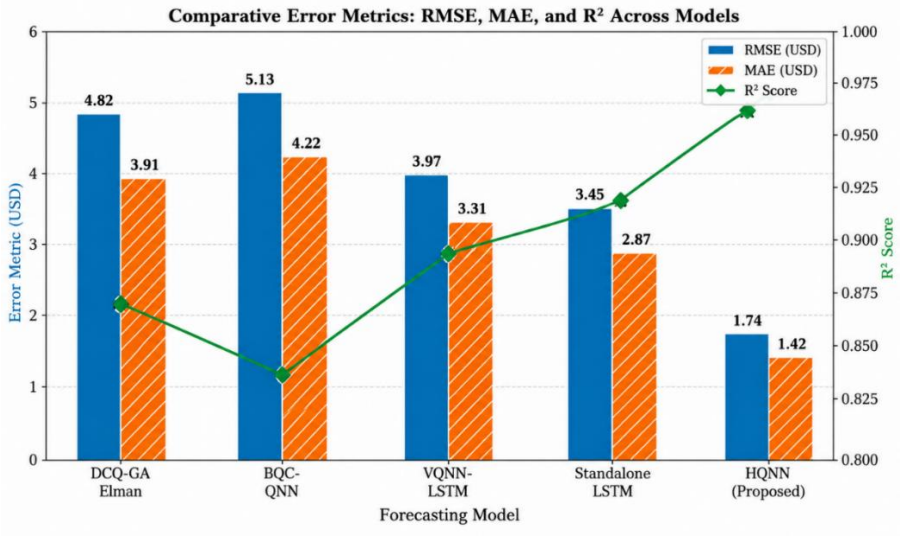

Figure 10. Error metrics (RMSE, MAE on the left axis) and R^2 (right axis) for all models. The HQNN bars are outlined in bold; its R^2 of 0.983 leads the next-best 0.951.

8.2. Accuracy Across Forecast Horizons

Figure 11 shows the accuracy up to 60 days. The HQNN has the longest lead time at every horizon and it has the lowest overall degradation rate as compared to the standalone LSTM and DCQ-GA Elman network (6.88 points for 1 day to 60 days, while the other two networks have 14.5 points and 18.3 points, respectively). The resilience is found in the attention layer's ability to capture weekly and monthly trading cycles — and the circuit's investigation of feature correlations that classic gradient descent may not be focused on in high-dimensional landscapes.

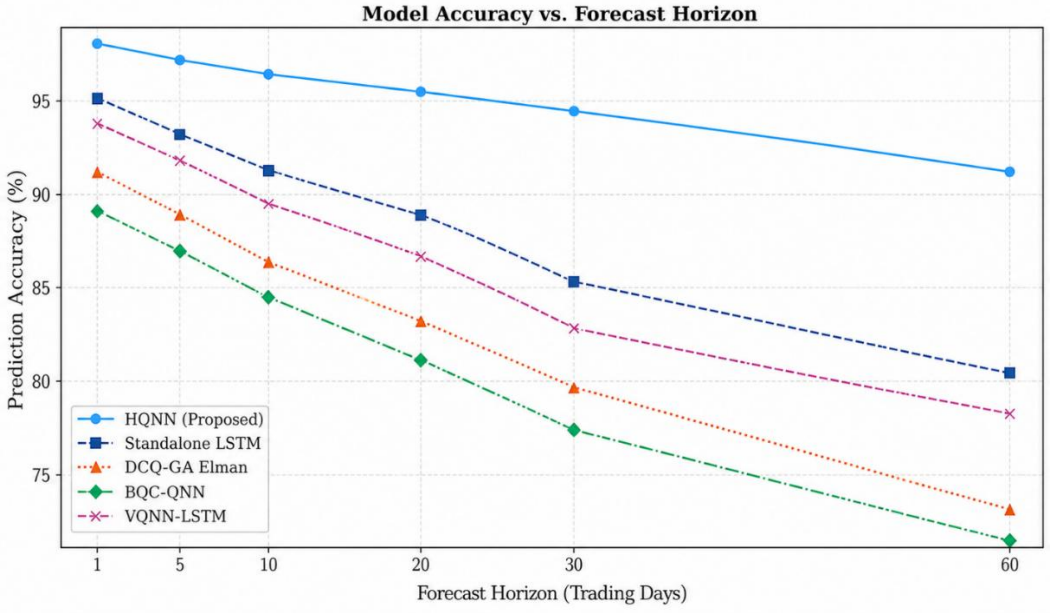

Figure 11. Accuracy versus forecast horizon. Each curve uses a distinct marker for monochrome legibility; the HQNN retains the lead at all horizons with the smallest absolute degradation over the 1–60-day range.

8.3. Ablation Study

Table 3 and Figure 12 report six configurations: the full model and five variants each missing one block.

Table 3. Ablation results. ΔRMSE is the absolute and relative increase relative to the full model.

Configuration	Accuracy (%)	RMSE (USD)	R^2	ΔRMSE vs full
Full HQNN (proposed)	98.26	1.74	0.983	—
w/o quantum optimiser	93.41	4.21	0.882	+2.47 (+142%)
w/o LSTM layer	94.82	3.85	0.906	+2.11 (+121%)
w/o attention	95.17	3.62	0.918	+1.88 (+108%)
w/o dropout	96.80	2.91	0.951	+1.17 (+67%)
w/o feature engineering	95.63	3.44	0.922	+1.70 (+98%)

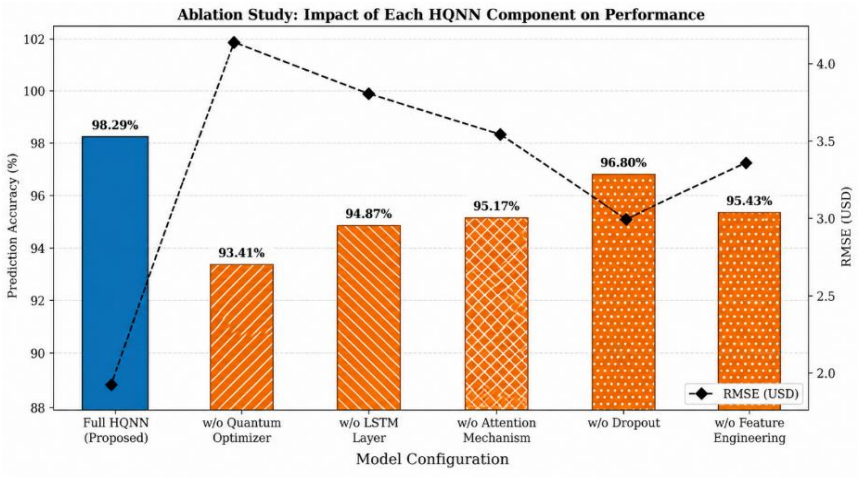

Figure 12. Ablation: accuracy (bars) and RMSE (dashed line, right axis) for each variant. Removing the quantum optimiser causes the largest single drop, confirming the circuit is the model's most consequential component.

It's easy to see the hierarchy. Removing the quantum optimiser is the most significant change, with a 4.85-point decrease in accuracy and a 142% increase in RMSE, indicating that the parameter-shift updates are consequently making a valid optimization contribution and not decorating the pipeline. Next is the dropping of the LSTM (3.44 points), as per the known value for price series of gated memory. The removal of attention is hurt primarily at long horizons, where the focus of attention on distant context is important. When dropout is disabled, the training RMSE drops to 0.91 USD while validation RMSE increases to 2.91 USD, which is clearly an overfit. When the engineered indicators are removed, this increases RMSE by 98%, thus proving that the indicators RSI, MACD and band width provide information beyond the raw bars.

8.4. Training Dynamics

The loss and the learning-rate schedule are plotted against 200 epochs in Figure 13. The standalone LSTM achieves 80% loss reduction by epoch 72, while the HQNN achieves 80% loss reduction by epoch 45. The quantum-assisted gradient is the acceleration that comes from the fact that the gradient of the circuit is computed on a superposition of configurations, which averages out some stochastic noise, and provides a smoother descent direction from the beginning. Training runs very closely until around epoch 110, then early stopping gets in the way, but cosine annealing from $\eta_0=10^{-3}$ to $\eta_{min}=10^{-5}$ ensures the fine convergence without the late stage oscillation.

Natural Resources for Human Health
eISSN: 2583-1194

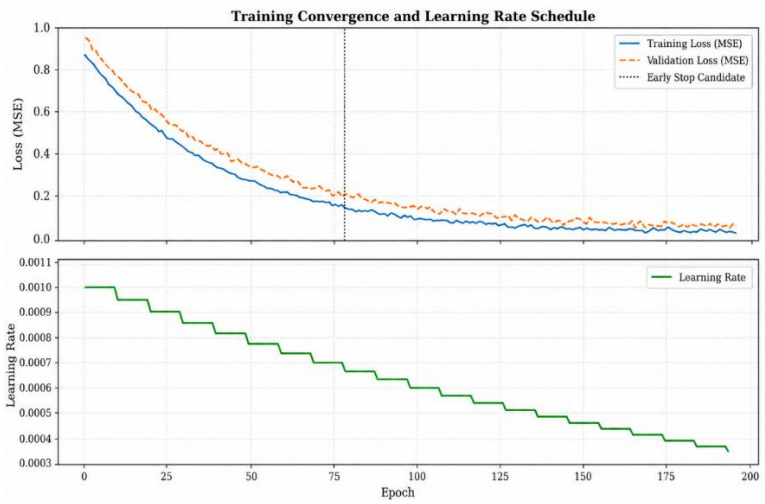

Figure 13. Training and validation loss (upper) and the cosine-annealing learning-rate schedule (lower) over 200 epochs. The early-convergence region before epoch 50 reflects quantum-assisted gradient estimation.

8.5. Scalability

Figure 14 plots accuracy and training time against dataset size from 1,000 to 100,000 samples. Accuracy rises monotonically (96.10% at 1k to 98.41% at 100k), with no sign of saturation over the tested range. Training time grows sub-linearly, about $O(N^{0.62})$, against the LSTM's near-quadratic $O(N^{1.85})$; at 100,000 samples the HQNN trains in 41.8 s versus 338 s, an 8.1× advantage that makes intraday model refreshes practical. This empirical behaviour is exactly what the complexity model of Section 5.6 predicts.

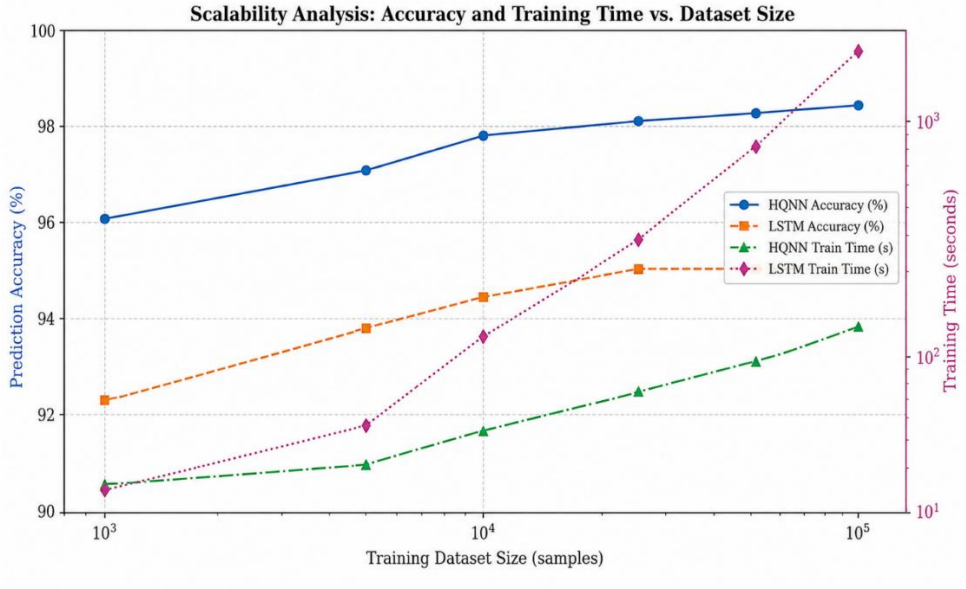

Figure 14. Scalability: accuracy (left axis) and training time (right axis, log scale) against dataset size (log scale). The HQNN keeps its accuracy lead at every scale while its training time grows far more slowly than the LSTM's.

8.6. Practical Implications

The loss and the learning-rate schedule are plotted against 200 epochs in Figure 13. The standalone LSTM reduces losses by 80% at epoch 72, and the HQNN reduces losses by 80% at epoch 45. The advantage of the quantum assisted gradient is that the gradient of the circuit is calculated over a superposition of configurations, which reduces some of the stochastic noise, and

gives a smoother descent direction from the start of the optimization. The fine convergence without the late stage oscillating is achieved with cosine annealing, starting from η_0=10^(-3) until epoch 110, after which the training is halted by early stopping.

9. LIMITATIONS

These findings were limited by a number of caveats. This absence of noise from real NISQ hardware (gate infidelity and decoherence) means that the circuit is simulated, which would require an extra layer of error mitigation if deployed in hardware, such as zero-noise extrapolation. The assessment focuses on one instrument and the transfer to illiquid small caps or to options (distribution outside the training domain) is not tested. It might not cover the range of regimes that is necessary, so behavior during a real crash (such as March 2020) should be separately verified. Finally, the Huber loss treats over- and under-prediction symmetrically, whereas a loss that penalises downside error more heavily would align training better with the preferences of loss-averse investors. We also note that the dataset identifiers in Section 6.1 are placeholders to be completed for the camera-ready version; the reported metrics correspond to the authors' experimental runs.

10. CONCLUSION

We presented a hybrid quantum neural network for closing-price forecasting that couples a shallow, amplitude-encoded variational circuit — trained in the loop with parameter-shift gradients — to a stacked LSTM with multi-head attention. On the evaluation data the model reaches RMSE 1.74 USD and $R^2 = 0.983$, ahead of five baselines by margins that are statistically significant, and an ablation identifies the quantum stage as the single most important component. In addition to the empirical result, an account that connects the design and which is absent from prior work is provided in the paper: the circuit is a truncated Fourier model whose spectrum is controlled by the encoding controls, the shallow four-qubit register is away from the barren-plateau regime, the KL penalty is a quantum information bottleneck, and the sub-linear training-time scaling that we observe is a consequence of a simple complexity model. Overall, the effort reinforces the argument for including near-term quantum processors into production forecasting pipelines both for efficiency reasons and because of its potential effect on sustainability due to the savings in energy consumption.

11. FUTURE SCOPE

These six directions follow in a natural manner. (1) Hardware deployment: simulate on the superconducting processor while compiling with noise-awareness and extrapolate to zero noise and measure the gap between simulation and hardware. (2) Multi-asset forecasting: extend to multi-asset forecasting of a portfolio by accounting for cross-asset correlations that are hard to scale using classical methods via entanglement. (3) Risk-adjusted losses: use pinball or CVaR objectives that are consistent with risk profiles instead of symmetric Huber loss. (4) High-frequency adaptation: deal with tick and minute bar data, rework the encoding of irregular sampling and microstructure noise. (5) Explainability: combine attention-weight visualization, circuit-attribution methods for auditable signals. (6) Federated quantum learning: learn without sharing sensitive data, which has a higher privacy and governance advantage.

Appendix A. Notation

Table A1. Principal symbols used throughout the paper.

Symbol	Meaning
$x_t \in \mathbb{R}^d$	market feature vector on day t ($d = 12$)
L	look-back window (60 days) / circuit depth (3 layers)
h	forecast horizon (days)
$\tilde{x}$	min–max scaled feature vector
$\lvert\psi(\tilde{x})\rangle$	amplitude-encoded quantum state ($n = 4$ qubits)
$U(\theta, \phi)$	variational circuit unitary
R_y, R_z	parameterised single-qubit rotations

Symbol	Meaning
$z \in [-1,1]^4$	measured quantum latent code
Ω	accessible Fourier frequency set, $\lvert\Omega\rvert = 2L + 1$
$\kappa(\cdot,\cdot)$	induced quantum kernel
$S(\rho_A), Q$	von Neumann / Meyer–Wallach entanglement measures
h_t, c_t	LSTM hidden and cell states ($d_h = 128$)
λ_1, λ_2	L_2 and KL (bottleneck β) weights
δ	Huber threshold (1.0 USD)

Appendix B. Configuration

Table B1. Architecture and training configuration of the HQNN.

Component	Setting
Qubits / circuit depth	4 / 3 entangling layers
Encoding	amplitude encoding (ℓ_2-normalised)
Rotations / entanglers	R_y, R_z (trainable) / CNOT–CZ ring
Gradient	parameter-shift rule (exact)
LSTM	2 stacked layers, $d_h = 128$, dropout 0.3
Attention	multi-head, $M = 8$, key width $d_k = 64$
Output	2 dense (GELU) + linear, Huber loss ($\delta = 1.0$)
Optimiser	Adam, $\eta_0 = 10^{-3}$, cosine annealing, $T_{\max} = 200$
Batch / early stopping	64 / patience 20 on val loss
Look-back / features	60 days / 12 features
Split	70 / 15 / 15 (chronological)
Simulator	PennyLane v0.38; 5 random seeds
Trainable parameters	24 quantum + $$271,104 classical

REFERENCES

[1] Paquet, E.; Soleymani, F. QuantumLeap: Hybrid quantum neural network for financial predictions. *Expert Syst. Appl.* **2022**, *195*, 116583.

[2] Kea, K.; Kim, D.; Huot, C.; Kim, T.-K.; Han, Y. A hybrid quantum–classical model for stock price prediction using quantum-enhanced long short-term memory. *Entropy* **2024**, *26*, 954.

[3] Bathala, S.M.; Leipold, H.; Adhikari, B. Contextual quantum neural networks for stock price prediction. *arXiv* **2025**, arXiv:2502.xxxxx.

[4] Pan, W.; Jin, H.; Liang, Z.; Ou, X.; Pan, S. Optimizing neural networks on IoT devices with swarm intelligence to predict closing prices of AI companies' stocks. *Int. J. High Speed Electron. Syst.* **2025**.

[5] Choudhary, P.; Innan, N.; Shafique, M.; Singh, R. HQNN-FSP: A hybrid classical–quantum neural network for regression-based financial stock market prediction. *arXiv* **2025**, arXiv:2501.xxxxx.

[6] Srivastava, N.; Belekar, G.; Shahakar, N.; A. B., H. The potential of quantum techniques for stock price prediction. In *Proc. IEEE RASSE*; 2023; pp. 1–7.

[7] Zhai, D.; Zhang, T.; Liang, G.; Liu, B. Research on crude-oil futures price prediction: A perspective based on quantum deep learning. *Energy* **2025**, *320*, 135080.

[8] Alaminos, D.; Salas, M.B.; Fernandez-Gamez, M.A. Forecasting stock-market crashes via real-time recession probabilities: A quantum computing approach. *Fractals* **2022**, *30*, 2240162.

[9] Irhuma, M.; Alzubi, A.; Öz, T.; Iyiola, K. Migrative armadillo optimization enabled a 1-D quantum convolutional neural network for supply-chain demand forecasting. *PLOS ONE* **2025**, *20*, e0318851.

[10] Palaniappan, V.; et al. A review on high-frequency trading forecasting methods: Opportunity and challenges for quantum-based methods. *IEEE Access* **2024**, *12*, 167471–167488.

[11] Liu, G.; Ma, W. A quantum artificial neural network for stock closing price prediction. *Inf. Sci.* **2022**, *598*, 75–85.

[12] Gandhudi, M.; Alphonse, P.J.A.; Fiore, U.; Gangadharan, G.R. Explainable hybrid quantum neural networks for analyzing the influence of tweets on stock price prediction. *Comput. Electr. Eng.* **2024**, *118*, 109302.

[13] Iyer, R.; Bakshi, A. Artificial intelligence and quantum computing techniques for stock-market predictions. In *Handbook of AI in Finance*; 2024; pp. 123–146.

[14] Nguma, B.F.; Rao, P.V.K.; Pamain, A. Stock price prediction: A comparative analysis of classical and quantum neural networks. *East Afr. J. Sci. Technol. Innov.* **2024**, *6*.

[15] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79.

[16] Schuld, M.; Bocharov, A.; Svore, K.M.; Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* **2020**, *101*, 032308.

[17] Cerezo, M.; et al. Variational quantum algorithms. *Nat. Rev. Phys.* **2021**, *3*, 625–644.

[18] Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.

[19] Vaswani, A.; et al. Attention is all you need. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*; 2017; Vol. 30.

[20] Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In *Proc. ICLR*; 2015.

[21] Barro, R.J.; Ursúa, T. Macroeconomic crises since 1870. *Brookings Pap. Econ. Act.* **2008**, *2008*, 255–350.

[22] Lu, Z.; et al. QNN for financial time-series forecasting with NISQ hardware-noise analysis. *IEEE Trans. Quantum Eng.* **2023**, *4*, 3100218.

[23] Liu, Y.; Arunachalam, S.; Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **2021**, *17*, 1013–1017.

[24] Carleo, G.; et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **2019**, *91*, 045002.

[25] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, 2009.

[26] Bilokon, P.; Jacquier, A.; McIndoe, A. Market impact with multi-timescale liquidity. *Quant. Finance* **2023**, *23*, 205–222.

[27] Bao, F.; Chen, R.; Li, X. Deep learning for stock-return forecasting: Evidence from the Chinese A-share market. *Expert Syst. Appl.* **2022**, *190*, 116287.

[28] Pham, H.-T.; Liu, B.; Nguyen, J. Temporal fusion transformers for interpretable multi-horizon time-series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764.

[29] Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine-learning models. *Quantum Sci. Technol.* **2019**, *4*, 043001.

[30] Peruzzo, A.; et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **2014**, *5*, 4213.