Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Hybrid Quantum-Classical Deep Learning Model for Global Cotton Yield Forecasting

Pooja Purohit[1], Prachi Janrao[2]

Department of Artificial Intelligence and Data Science, Thakur College of Engineering and Technology, Mumbai, India.

E-mail: [1]purohitpooja321@gmail.com, [2]prachi.janrao@thakureducation.org

Orcid ID: [1]0009-0002-2726-6602, [2]0000-0002-1270-4254

Abstract

Global cotton yield forecasting is critical for food security planning, supply chain management, and climate adaptation policy, yet remains challenging due to the non-linear interactions among satellite vegetation indices, meteorological variables, and agronomic factors across heterogeneous growing regions. This paper proposes the Hybrid Quantum-Classical Deep Learning (HQC-DL) framework, the first systematic empirical study integrating parametric quantum circuit (PQC)-based feature enhancement into a CNN-BiLSTM and gradient-boosted regression pipeline for global crop yield prediction. When tested on the FAOSTAT-MODIS-CRU dataset with 1,142 country-year observations of 59 countries between 2001 and 2020, the proposed method exhibits RMSE of 0.769 ± 0.004 t/ha and R^2 of 0.690 ± 0.004, and is characterized by the lowest cross-run variance compared to all other neural network architectures, which demonstrates its better prediction stability. The ablation study shows that all architectural stages have a positive contribution: when gradient boosted regressors are trained on features generated via deep learning methods, they show better results than models trained on raw data, regardless of the quantum module, and PQC expressibility analysis proves that the ansatz is capable of exploring a wide range of states in Hilbert space.

Keywords: Quantum Machine Learning, Parametric Quantum Circuit, Hybrid Quantum-Classical, Cotton Yield Prediction, CNN-BiLSTM, XGBoost, NDVI, FAOSTAT, Crop Forecasting.

1. Introduction

Food security around the globe has become one of the most critical problems of the current century, with the world's population expected to exceed 9.7 billion by 2050. Cotton (Gossypium hirsutum) is the most important natural fiber for textiles on the earth, grown in a variety of climatic zones around the world over an area of 35 million hectares. Developing a critical prediction of cotton yield in a timely manner is essential for policy decisions, supply chain optimization, and early warning systems against shortages due to climate variability, drought, or pest infestations.

Crop yield prediction is a very complex spatiotemporal prediction task controlled by the non-linear interaction of satellite-based vegetation indices, meteorological variables, and agronomic variables. A CNN-BiLSTM architecture was designed in the preceding work which this paper builds on for forecasting global cotton yields, demonstrating performance surpassing baseline models such as standalone LSTM, BiLSTM, SVM, and XGBoost. Nonetheless, a

DOI: https://doi.org/10.36548/jtcsst.2026.3.019
Received: 04.06.2026, received in revised form: 08.07.2026, accepted: 01.08.2026, published: 14.08.2026

systematic inspection of its constraints identified five critical shortcomings:

1. Constriction to a classical feature space with a representational capability of polynomial complexity.
2. Lack of complete cross-variable interdependency representation among NDVI, temperature, and precipitation.
3. Incapacity to discover latent high-dimensional relationships across multi-source data.
4. Insufficient generalisation across geographically varied regions and inter-annual seasonal variations.
5. Absence of a hybrid feature enhancement mechanism.

Quantum machine learning (QML) and Parametric Quantum Circuits (PQCs) operate in a 2^n-dimensional Hilbert space for n qubits. This gives an exponentially larger feature space than any classical analogue, allowing the encoding of non-linear, complex feature relationships through quantum superposition and entanglement [2]. PQCs can be trained by the parameter-shift rule [3, 4] and integrated with classical deep learning frameworks through PennyLane [5]. Based on these observations, this paper proposes the HQC-DL framework to predict global cotton yield. The main contributions are as follows:

1. We introduce the first systematic empirical study of PQC-based feature enhancement with a CNN-BiLSTM-XGBoost pipeline to predict crop yields, with reproducible benchmarking results across eight model variants.
2. We present a concatenated classical-quantum feature architecture that retains the entire 256-dimensional BiLSTM representation alongside 6-dimensional PQC expectation values, overcoming the information bottleneck of compressive quantum encoding and demonstrating improved prediction stability (lowest cross-run variance among all neural network-based models).
3. We show through ablation experiments that gradient-boosted regressors applied to deep learning-generated features consistently outperform the same regressors on raw engineered features, independently of the quantum component.
4. We tested the framework on a globally representative FAOSTAT-MODIS-CRU dataset in 59 cotton-producing countries.
5. We report PQC expressibility metrics that provide empirical evidence that the ansatz explores a meaningfully large region of Hilbert space.

2. Literature Review

2.1 Traditional and Statistical Approaches

The traditional technique of forecasting crop yield was based on using process-based simulation models like DSSAT, APSIM, and WOFOST [6]. Such models mechanically incorporate the physiology of crops along with soil dynamics; however, they involve extensive region-based parameters, thus making them difficult to scale up globally due to differences in growing conditions. The statistical method involves the use of time series models belonging to the ARIMA class of models [7]; however, such methods rely on linearity and stationarity, assumptions that get violated more often than not.

2.2 Machine Learning Approaches

Ensemble and kernel-based approaches to machine learning mitigated some of the limitations associated with statistical models. The Random Forest algorithm showed promise as a reliable non-linear approach in a variety of benchmark crop yield forecasts, whereas the Support Vector Regression method with RBF kernels, although quite effective for small samples, is not very scalable when it comes to multi-source big data [14]. The XGBoost framework turned out to be an effective solution for agricultural tabular data prediction due to its gradient boosting regularization and flexibility for working with feature spaces of different natures [8]. The results of a systematic review of 50 crop yield forecast studies [9] have shown that ensemble tree methods and deep learning models contribute in the same way to achieving the best pipeline performance.

2.3 Deep Learning for Crop Yield Prediction

The deep learning models have advanced crop yield prediction beyond the limitations of feature engineering in classical machine learning techniques. The LSTM model showed an ability to learn the temporal dependencies in seasonal agricultural time series [15], and the BiLSTM improved this ability by making the learning bidirectional and hence more sensitive to phenological patterns exhibited differently in forward and backward time directions [16]. The combination of CNN and LSTM allowed the modeling of both spatial and temporal features from gridded input data [18], while recurrent neural networks have been used for predicting crop yield parameters through LRNN models [17]. Recently, transformers have been explored for forecasting agricultural time series data, but their advantages over CNN-BiLSTM for moderate-sized datasets are not consistent, especially when there is little training data compared to the model capacity. In all these models, there is one common limitation: the learned representation is always constrained within classical Euclidean feature spaces.

2.4 Remote Sensing and Multi-Source Data Integration

The existence of satellite-based vegetation indices and globally harmonized reanalysis climate products has significantly increased the number of features used in crop prediction models. NDVI estimated from MODIS [19] data can be considered a well-researched predictor of crops' biomass and development phases, showing the dynamics of canopy greening throughout the seasons on a global scale. Yield statistics collected by FAO and harmonized by the FAOSTAT [1] project act as a standard target variable for all nations and periods, whereas meteorological reanalysis data like CRU TS4.08, available through the World Bank Climate Change Knowledge Portal, act as predictors of temperature and precipitation. It should be noted that merging the data from these three sources within a single predictive model is not a trivial task, since the predictors vary in terms of their spatial and temporal resolutions and distributions.

2.5 Quantum Machine Learning

The parametric quantum circuits provide a new paradigm of feature encoding. PQCs map classical inputs into quantum states using angle embedding, transform these states using parameterised unitary maps for quantum feature enhancement and output classical values by measuring expectations [20]. They are differentiable through the parameter shift rule [3, 4] and can be seamlessly incorporated into popular deep learning software using PennyLane [5]. Efficient hardware-efficient circuit architectures trade off expressiveness and depth, which is an

important design choice within the NISQ paradigm [21]. The expressibility score [10] quantifies the uniformity of sampling from the unitary group by a particular ansatz and allows for the proper evaluation of the circuit architecture irrespective of its performance on any particular task.

2.6 Hybrid Quantum-Classical Models

Due to limitations in existing quantum devices, hybrid quantum-classical approaches have become the prevailing methodology for near-term applications of quantum machine learning [21]. Hybrid approaches exploit quantum representation properties of the quantum circuit by utilizing classical pre- and post-processing in conventional neural networks. Positive outcomes have been achieved for tasks such as image classification [22], natural language processing, and financial time series prediction [20]. Nevertheless, agricultural yield prediction using the hybrid quantum-classical approach has not yet been investigated, and no study has examined the practicality of quantum representation properties for this problem.

2.7 Research Gap

The proposed framework bridges three gaps, the combination of which constitutes the motivation behind this study. Current deep learning models for crop forecasting are restricted to classical feature sets and are unable to benefit from the representation provided by superposition and entanglement. At the same time, PQCs provide a theoretically grounded and empirically tractable mechanism for generating quantum-enhanced features that are differentiable and integrable with existing deep learning pipelines, and gradient-boosted ensemble methods offer complementary strengths in robustness and generalisation that are not fully exploited when applied to raw inputs rather than to deep-learned or quantum-enhanced representations. The HQC-DL framework proposed in this paper is the first to systematically integrate all three paradigms in a reproducible, multi-source, global-scale agricultural forecasting pipeline, directly addressing each of these gaps.

3. Methodology

3.1 Overall Framework

The HQC-DL framework is structured as a three-stage sequential pipeline illustrated in Figure 1. Let $\mathbf{X} \in \mathbb{R}^{T \times F}$ denote the input matrix for a single country-year observation. The overall model is:

$$\hat{y} = f_{\text{XGB}}\left(\left[\, \mathcal{M}_{\boldsymbol{\theta}}^{\text{PQC}}(\mathbf{h}) \parallel \mathbf{h} \,\right]\right), \quad \mathbf{h} = f_{\text{BiLSTM}}(f_{\text{CNN}}(\mathbf{X}; \mathbf{W}_c); \mathbf{W}_b) \tag{1}$$

where $\mathbf{h} \in \mathbb{R}^{256}$ is the BiLSTM output, $\mathcal{M}_{\boldsymbol{\theta}}^{\text{PQC}}(\mathbf{h}) \in [-1, 1]^6$ is the PQC expectation value vector, $\parallel$ denotes vector concatenation yielding a 262-dimensional combined representation, and f_{XGB} is the gradient-boosted regressor.

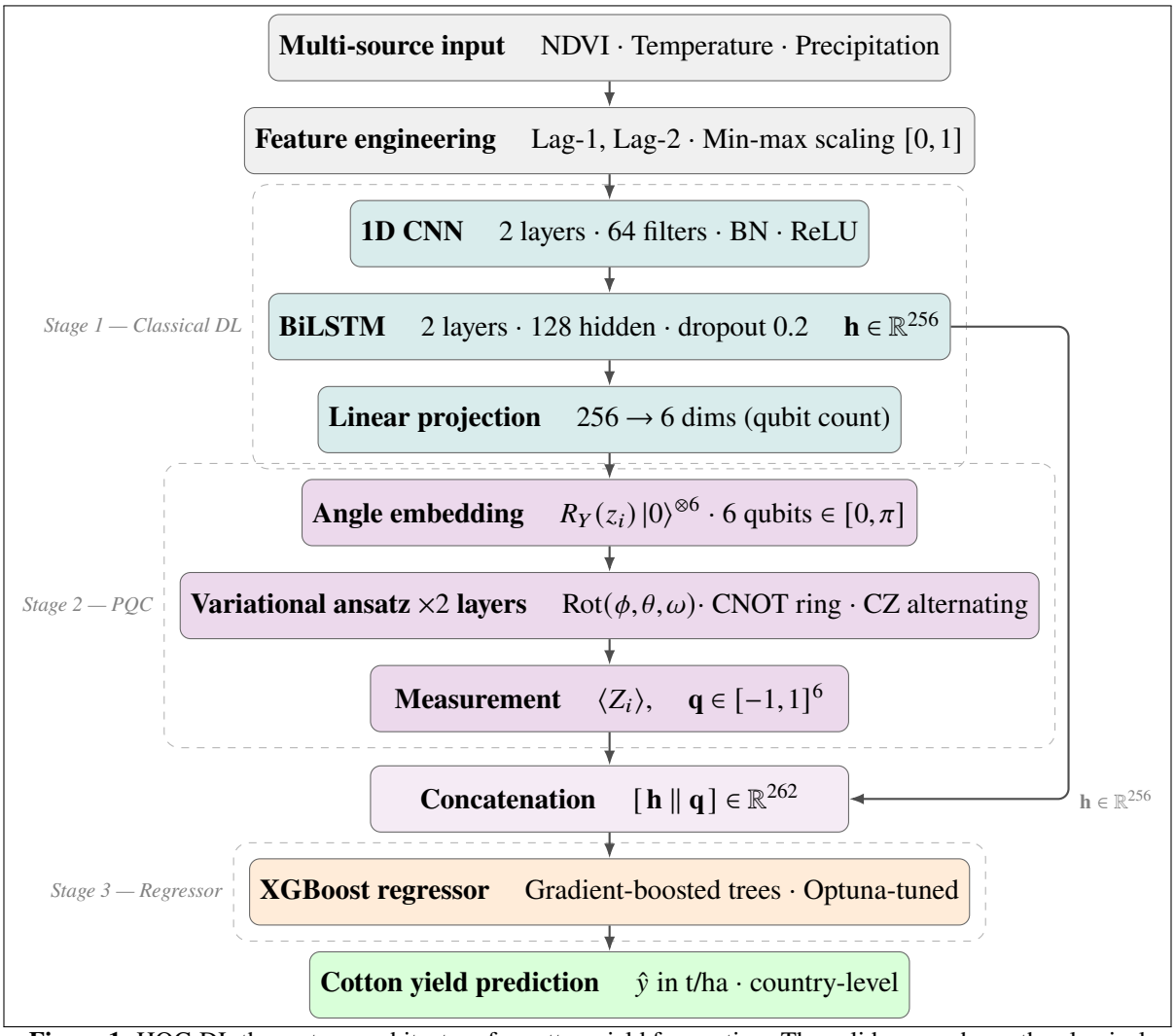

Figure 1. HQC-DL three-stage architecture for cotton yield forecasting. The solid arrow shows the classical BiLSTM feature vector **h** bypassing the PQC and joining the quantum output at the concatenation layer before XGBoost

Algorithm formalises the complete two-phase training and inference procedure implementing Eq.(1).

Algorithm 1 HQC-DL: Hybrid quantum-classical Deep learning for cotton yield forecasting

Require: Multi-source dataset $\mathcal{D} = \{(\mathbf{X}_i, y_i)\}_{i=1}^{N}$; chronological split (train 2001–2015, val 2016–2017, test 2018–2020); PQC qubits $n = 6$, ansatz layers $D = 2$; max epochs $E = 100$, patience $P = 15$, learning rate $\eta = 5 \times 10^{-4}$

Ensure: Predicted cotton yield $\hat{y}$ in t/ha

// Preprocessing

Construct lag-1 and lag-2 features: $x^{(t-k)} = x_{t-k},\ k \in \{1, 2\}$ ▷ 45 features total

Fit min-max scaler on training partition; apply to all splits

// Phase 1: Joint end-to-end training of CNN-BiLSTM-PQC

Initialise CNN weights $\mathbf{W}_c$, BiLSTM weights $\mathbf{W}_b$, projection weights $\mathbf{W}_{\text{proj}}$, PQC parameters $\boldsymbol{\theta}$, FC head weights $\mathbf{W}_{\text{fc}}$

for epoch = 1 to E **do**

  for each mini-batch $(\mathbf{X}, y) \in \mathcal{D}_{\text{train}}$ **do**

    $\mathbf{H}^{\text{conv}} \leftarrow \text{ReLU}(\text{BN}(\mathbf{W}_c * \mathbf{X}))$ ▷ 1D CNN

    $\mathbf{h} \leftarrow \text{BiLSTM}(\mathbf{H}^{\text{conv}}; \mathbf{W}_b), \quad \mathbf{h} \in \mathbb{R}^{256}$ ▷ BiLSTM

    $\mathbf{z} \leftarrow \pi \cdot \sigma(\mathbf{W}_{\text{proj}}\mathbf{h} + \mathbf{b}_{\text{proj}}), \quad \mathbf{z} \in [0, \pi]^6$ ▷ Linear projection

    $|\psi_{\text{enc}}\rangle \leftarrow \bigotimes_{i=1}^{n} R_Y(z_i)\,|0\rangle^{\otimes n}$ ▷ Angle embedding

    $|\psi_{\text{out}}\rangle \leftarrow U(\boldsymbol{\theta})\,|\psi_{\text{enc}}\rangle$ ▷ Variational ansatz (Rot + CNOT ring + CZ)

    $\mathbf{q} \leftarrow [\langle Z_i \rangle]_{i=1}^{n}, \quad \mathbf{q} \in [-1, 1]^6$ ▷ Pauli-Z measurement

    $\hat{y}^{\text{FC}} \leftarrow \mathbf{W}_{\text{fc}}\,\mathbf{q} + \mathbf{b}_{\text{fc}}$ ▷ Temporary FC head

    $\mathcal{L} \leftarrow \frac{1}{B}\sum(y - \hat{y}^{\text{FC}})^2$

    Backpropagate via parameter-shift rule through $\boldsymbol{\theta}$; autograd through $\mathbf{W}_{\text{proj}}, \mathbf{W}_b, \mathbf{W}_c$

    Update all parameters with Adam ($\eta = 5 \times 10^{-4}$, decay $\gamma = 0.95$)

  end for

  if validation MSE not improved for P consecutive epochs **then**

    break ▷ Early stopping

  end if

end for

Discard FC head; freeze $\mathbf{W}_c$, $\mathbf{W}_b$, $\mathbf{W}_{\text{proj}}$, $\boldsymbol{\theta}$

// Phase 2: XGBoost fitting on extracted features

for each observation $\mathbf{X}_i$ in train, val, and test **do**

  Extract $\mathbf{h}_i$ and $\mathbf{q}_i$ using the frozen pipeline (CNN through Pauli-Z measurement, as in Phase 1)

  $\mathbf{v}_i \leftarrow [\mathbf{h}_i \,\|\, \mathbf{q}_i] \in \mathbb{R}^{262}$ ▷ Concatenate classical and quantum features

end for

Tune XGBoost hyperparameters via Optuna (50 TPE trials) on $\mathcal{D}_{\text{val}}$

Fit XGBoost regressor f_{XGB} on $\{\mathbf{v}_i, y_i\}_{i \in \mathcal{D}_{\text{train}}}$ with tuned hyperparameters

// Inference

return $\hat{y}_i = f_{\text{XGB}}(\mathbf{v}_i)$ for each test observation $i \in \mathcal{D}_{\text{test}}$

3.2 Dataset Description

Cotton yield data was sourced from FAOSTAT [1]. Raw values in kg/ha were converted to t/ha by dividing by 1,000. The dataset spans 2001–2020 across 59 countries comprising 1,142 annual country-year observations. Yield statistics: mean 1.727 t/ha, std 1.187 t/ha, range $[0.068, 6.813]$ t/ha.

3.2.1 Satellite Vegetation Index Data

Monthly NDVI was obtained from MODIS MOD13A3 Collection 6.1 [19], accessed via Google Earth Engine. Country-level values were spatially aggregated using the USDOS LSIB boundary dataset. NDVI is defined as:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} \tag{2}$$

where ρ_{NIR} and ρ_{RED} are the near-infrared and red band surface reflectances, respectively. Six annual aggregate NDVI features were derived: mean, maximum, minimum, standard deviation, range, and peak month.

3.2.2 Meteorological Data

Monthly temperature and precipitation data were sourced from CRU TS4.08 at 0.5° resolution via the World Bank CCKP API. Nine annual aggregate climate features were derived including mean, maximum, minimum, and standard deviation of temperature; total, mean, standard deviation, maximum of precipitation; and count of months above 15°C (cotton germination threshold proxy). annual mean temperature was 21.51°C (range 1.90–29.76°C). and annual total precipitation was 1,062.2 mm (range 10.9–3,598.6 mm).

The distributions of these key variables across the dataset are visualized in Figure 2.

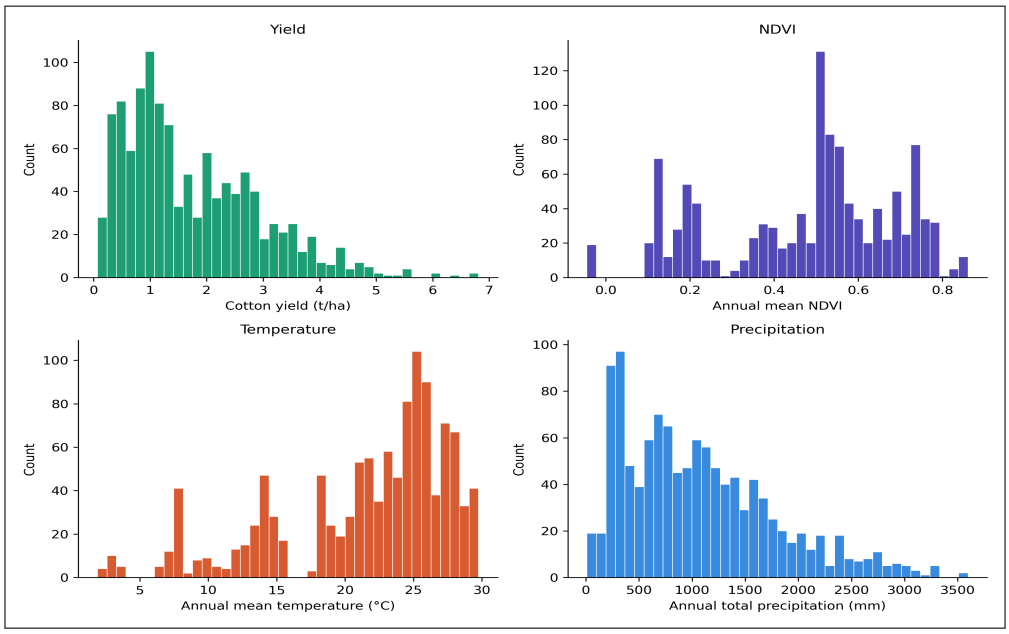

Figure 2. Dataset feature distributions. Clockwise from top-left: cotton yield (t/ha), annual mean NDVI, annual total precipitation (mm), and annual mean temperature (°C)

3.2.3 Final Feature Set

Table 1 summarizes the 15 base features (6 NDVI + 9 climate), expanded to 45 features after lag-1 and lag-2 augmentation.

Table 1. Input feature set

Feature	Source	Unit	Description
NDVI_mean	MODIS MOD13A3	—	Annual mean NDVI
NDVI_max	MODIS MOD13A3	—	Annual maximum NDVI
NDVI_min	MODIS MOD13A3	—	Annual minimum NDVI
NDVI_std	MODIS MOD13A3	—	Intra-annual NDVI variability
NDVI_range	MODIS MOD13A3	—	Annual NDVI range
NDVI_peak_month	MODIS MOD13A3	month	Month of peak NDVI
Temp_mean	CRU TS4.08	°C	Annual mean temperature
Temp_max	CRU TS4.08	°C	Annual maximum temperature
Temp_min	CRU TS4.08	°C	Annual minimum temperature
Temp_std	CRU TS4.08	°C	Temperature variability
Precip_total	CRU TS4.08	mm	Annual total precipitation
Precip_mean	CRU TS4.08	mm	Monthly mean precipitation
Precip_std	CRU TS4.08	mm	Precipitation variability
Precip_max	CRU TS4.08	mm	Maximum monthly precipitation
Months_above_15C	CRU TS4.08	count	Cotton growth season proxy
Yield (target)	FAOSTAT	t/ha	Cotton seed yield

3.3 Feature Engineering and Preprocessing

3.3.1 Temporal Lag Features

Lag features of order 1 and 2 are constructed for each variable x:

$$x^{(t-k)} = x_{t-k}, \quad k \in \{1, 2\} \tag{3}$$

yielding an augmented feature dimension $3F = 45$ per observation.

The extent of multi-year carryover effects on cotton yield comes from soil nutrient depletion, the presence of a perennial root system, seed quality from previous years' harvests, and multi-year rainfall/drought cycles [7]. Lag-1 features capture the immediate prior-year growing conditions that most directly influence current-year productivity through soil moisture carryover and crop calendar continuity. Lag-2 features capture biennial climate oscillations, particularly El Niño-Southern Oscillation (ENSO) cycles, which have a characteristic periodicity of 2–7 years and significantly affect cotton-growing regions in South Asia, West Africa, and Australia. Lag orders beyond $k = 2$ were excluded for two practical reasons: (i) each additional lag order removes the first k years per country from the training set (the first two years per country already become unavailable after $k = 2$ due to missing lag values at sequence boundaries, reducing 59 countries $\times$ 20 years = 1,180 potential observations to 1,142 usable observations); and (ii) autocorrelation analysis of the yield time series showed that partial autocorrelations at lags $k \geq 3$ were not statistically significant at the $\alpha = 0.05$ level across the majority of countries in the dataset.

3.3.2 Normalization

All features are normalised to $[0, 1]$ using min-max scaling fitted exclusively on the training partition:

$$\tilde{x} = \frac{x - x_{\min}^{\text{train}}}{x_{\max}^{\text{train}} - x_{\min}^{\text{train}}} \tag{4}$$

This is required for PQC angle embedding, which maps inputs to $[0, \pi]$ via multiplication by π.

3.3.3 Chronological Data Split

Train: 2001–2015 (842 obs), Val: 2016–2017 (120 obs), Test: 2018–2020 (180 obs)

3.4 Stage 1: CNN-BiLSTM Feature Extractor

3.4.1 1D Convolutional Neural Network

Given input $\mathbf{X} \in \mathbb{R}^{T \times F}$, the CNN applies two convolutional layers with batch normalisation and ReLU:

$$\mathbf{H}_l^{\text{conv}} = \text{ReLU}\big(\text{BN}\big(\mathbf{W}_l^{\text{conv}} * \mathbf{H}_{l-1}^{\text{conv}} + \mathbf{b}_l^{\text{conv}}\big)\big), \quad l = 1, 2 \tag{5}$$

The input tensor $\mathbf{X}$ has shape $(\text{batch} \times 1 \times F)$, where the sequence length is treated as 1 and $F = 45$ is the feature dimension. Both convolutional layers use 64 filters with kernel size $k = 3$ and `same` padding to preserve the feature dimension. Batch normalisation is applied after each convolution and before the ReLU activation to stabilise training. The CNN output $\mathbf{H}^{\text{conv}} \in \mathbb{R}^{\text{batch} \times F \times 64}$ is transposed to $(\text{batch} \times F \times 64)$ before being fed to the BiLSTM, treating the 64 convolutional channels as the per-timestep feature representation across the F feature positions. The total CNN parameter count is 12,864 parameters (measured; see Table 4).

3.4.2 Bidirectional LSTM

The BiLSTM cell computes:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \tag{6}$$
$$\mathbf{f}_t = \sigma\big(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f\big) \tag{7}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \tag{8}$$
$$\mathbf{g}_t = \tanh\big(\mathbf{W}_g[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_g\big) \tag{9}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \tag{10}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{11}$$

The bidirectional hidden state is:

$$\mathbf{h}_t^{\text{bi}} = \left[\overrightarrow{\mathbf{h}_t} \parallel \overleftarrow{\mathbf{h}_t}\right] \in \mathbb{R}^{2H} \tag{12}$$

The final state $\mathbf{h} = \mathbf{h}_T^{\text{bi}} \in \mathbb{R}^{256}$ is passed to Stage 2.

3.5 Stage 2: Parametric Quantum Circuit

3.5.1 Quantum State Encoding

The classical feature vector $\mathbf{h} \in \mathbb{R}^{256}$ is projected to qubit dimension $n = 6$ via a trainable linear layer:

$$\mathbf{z} = \mathbf{W}_{\text{proj}}\,\mathbf{h} + \mathbf{b}_{\text{proj}}, \quad \mathbf{z} \in \mathbb{R}^6 \tag{13}$$

The projected vector is scaled to $[0, \pi]$ via sigmoid and encoded using angle embedding:

$$|\psi_{\text{enc}}\rangle = \bigotimes_{i=1}^{n} R_Y(z_i)\,|0\rangle^{\otimes n} \tag{14}$$

where the Pauli-Y rotation gate is:

$$R_Y(\theta) = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \tag{15}$$

3.5.2 Hardware-Efficient Ansatz

Each ansatz layer $d \in \{1, 2\}$ applies the rotation, entanglement, and measurement sequence illustrated in Figure 3:

Rotation Sub-layer: A general single-qubit rotation:

$$\text{Rot}(\phi, \theta, \omega) = R_Z(\omega)\,R_Y(\theta)\,R_Z(\phi) = \begin{pmatrix} e^{-i(\phi+\omega)/2}\cos(\theta/2) & -e^{i(\phi-\omega)/2}\sin(\theta/2) \\ e^{-i(\phi-\omega)/2}\sin(\theta/2) & e^{i(\phi+\omega)/2}\cos(\theta/2) \end{pmatrix} \tag{16}$$

Entanglement Sub-layer: CNOT ring topology:

$$\text{CNOT}_{i \to (i+1) \bmod n}, \quad i = 0, 1, \ldots, n-1 \tag{17}$$

followed by CZ gates on alternating pairs:

$$\text{CZ}_{i \to i+1}, \quad i \in \{0, 2, 4\} \tag{18}$$

The full D-layer ansatz unitary is:

$$U(\boldsymbol{\theta}) = \prod_{d=1}^{D} \left[\mathcal{E}_{\text{CZ}} \cdot \mathcal{E}_{\text{CNOT}} \cdot \bigotimes_{i=1}^{n} \text{Rot}(\phi_i^{(d)}, \theta_i^{(d)}, \omega_i^{(d)}) \right] \tag{19}$$

Total trainable PQC parameters: $3nD = 3 \times 6 \times 2 = 36$.

The choice of $n = 6$ qubits reflects a balance of practical and theoretical considerations rather than a single deciding factor. The linear projection layer reduces the 256-dimensional BiLSTM output to n dimensions before encoding; $n = 6$ provides a $2^6 = 64$-dimensional Hilbert space that is exponentially larger than any 6-dimensional classical embedding, while remaining tractable on the default.qubit statevector simulator used in this study. This was further supported by the expressibility analysis, which confirmed that 6 qubits with 2 ansatz layers achieves a KL divergence close to zero relative to the Haar-random distribution, indicating sufficient exploration of the Hilbert space. Going beyond this, to 8 or more qubits, would increase simulation time super-exponentially (statevector memory scales as 2^n), making 3-seed cross-validation impractical on Colab hardware within the study constraints. The selection of $D = 2$ ansatz layers was based on the observation that a single layer is insufficient for full n-qubit entanglement through the ring CNOT topology (each qubit reaches all others in $\lceil n/2 \rceil = 3$ layers

for $n = 6$, but 2 layers provides transitivity via the CZ alternating pairs), while $D \geq 3$ risks *barren plateaus*, i.e., exponential gradient vanishing in parameterised quantum circuits, without additional techniques such as layer-wise training. Together, $n = 6$ and $D = 2$ yield 36 trainable PQC parameters, a compact circuit depth of $2(n + \lfloor n/2 \rfloor) = 18$ gates per layer, and expressibility confirmed empirically.

3.5.3 Quantum Measurement and Combined Output

The quantum output state is:

$$|\psi_{\text{out}}\rangle = U(\boldsymbol{\theta})\, |\psi_{\text{enc}}\rangle \tag{20}$$

Pauli-Z expectation values are measured on each qubit:

$$q_i^{\text{out}} = \langle \psi_{\text{out}}|\, Z_i\, |\psi_{\text{out}}\rangle, \quad i = 1, \ldots, n \tag{21}$$

yielding $\mathbf{q}^{\text{out}} \in [-1, 1]^6$. The combined 262-dimensional feature vector passed to XGBoost is:

$$\mathbf{v} = \left[\mathbf{h} \parallel \mathbf{q}^{\text{out}}\right] \in \mathbb{R}^{262} \tag{22}$$

3.5.4 Differentiability via the Parameter-Shift Rule

The PQC is trained via the parameter-shift rule [3, 4]:

$$\frac{\partial \langle O \rangle}{\partial \theta} = \frac{1}{2}\left[\langle O \rangle_{\theta+\pi/2} - \langle O \rangle_{\theta-\pi/2}\right] \tag{23}$$

3.5.5 Circuit Expressibility

Circuit expressibility [10] is quantified as the Kullback-Leibler divergence between the PQC fidelity distribution and the Haar-random unitary ensemble:

$$\mathcal{E} = D_{\text{KL}}\left(\hat{P}_{\text{PQC}}(F; \boldsymbol{\theta}) \parallel P_{\text{Haar}}(F)\right) \tag{24}$$

A smaller $\mathcal{E}$ indicates higher expressibility.

Encoding: $R_Y(z_i)$ applies angle embedding of the projected input $z_i \in [0, \pi]$ to qubit i.

Rot: $\text{Rot}(\phi_i^{(d)}, \theta_i^{(d)}, \omega_i^{(d)}) = R_Z(\omega) R_Y(\theta) R_Z(\phi)$, one per qubit per layer.

CNOT Ring: $q_1 \to q_2 \to q_3 \to q_4 \to q_5 \to q_6 \to q_1$, shown above for one ansatz layer as six sequential CNOT gates, with the final $q_1 \to q_6$ wraparound gate (rightmost column before measurement) completing the ring. The second ansatz layer ($D = 2$) repeats this identical Rot + CNOT ring + CZ structure with independently trained parameters $\phi_i^{(2)}, \theta_i^{(2)}, \omega_i^{(2)}$ and is omitted here for clarity. CZ gates on alternating pairs $(q_1, q_2), (q_3, q_4), (q_5, q_6)$ are applied after the CNOT ring in each layer (omitted from this diagram; see Eq.(19) for the complete unitary).

Measurement: Pauli-Z expectation $\langle Z_i \rangle$ on each qubit yields $\mathbf{q} \in [-1, 1]^6$.

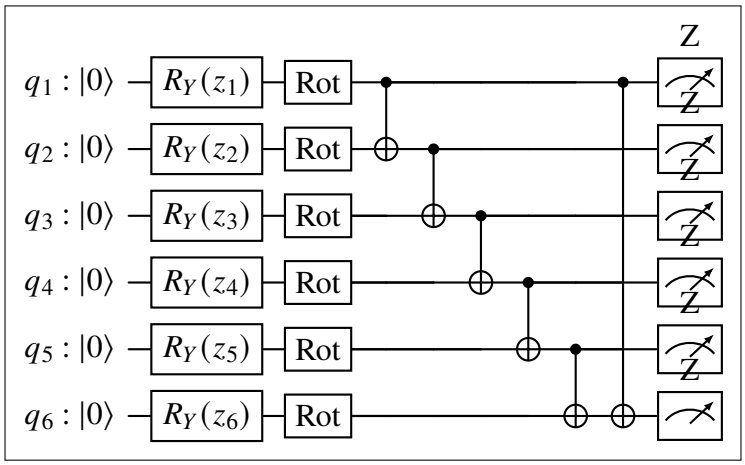

Figure 3. PQC architecture for one representative ansatz layer: angle embedding, Rot gates, and full CNOT ring entanglement (including the $q_1 \to q_6$ wraparound, shown as a separate gate following the five sequential forward CNOTs). The full circuit applies this structure across $D = 2$ layers, each followed by a CZ sub-layer on alternating qubit pairs, before final measurement. Total trainable parameters: $3 \times 6 \times 2 = 36$.

3.6 Stage 3: XGBoost Regressor

The 262-dimensional vector $\mathbf{v}$ is passed to an XGBoost gradient-boosted regressor [8]:

$$\hat{y} = \sum_{k=1}^{K} f_k(\mathbf{v}), \quad f_k \in \mathcal{F} \tag{25}$$

trained by minimising the regularised objective:

$$\mathcal{L}_{\text{XGB}} = \sum_{i=1}^{N} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \left(\gamma T_k + \tfrac{1}{2} \lambda \|\mathbf{w}_k\|^2 \right) \tag{26}$$

Hyperparameters are optimised using Optuna [11] with 50 TPE trials on the validation set.

3.7 Training Procedure

The CNN-BiLSTM and PQC components are trained jointly using the Adam optimiser [25] with learning rate 5×10^{-4} and exponential decay ($\gamma = 0.95$). Training objective:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} (y_i - \hat{y}_i)^2 \tag{27}$$

Training proceeds in two phases. In the first phase, the CNN-BiLSTM and PQC components are optimised jointly in an end-to-end fashion using a temporary fully connected (FC) regression head appended after the PQC measurement layer. The MSE loss is computed between the FC output $\hat{y}_i^{\text{FC}}$ and the ground truth yield y_i, and gradients are backpropagated through the FC head, the PQC parameters $\boldsymbol{\theta}$ via the parameter-shift rule, the projection layer $\mathbf{W}_{\text{proj}}$, the BiLSTM weights $\mathbf{W}_b$, and the CNN weights $\mathbf{W}_c$ in a single unified optimisation step using PyTorch's `autograd` engine and PennyLane's `TorchLayer` interface. In the second phase, after convergence (determined by early stopping on validation MSE), the FC head is discarded. The frozen CNN-BiLSTM-PQC pipeline is used as a fixed feature extractor to

produce the 262-dimensional concatenated classical-quantum feature vector $\mathbf{v}$ for all training, validation, and test observations. XGBoost is then fitted independently on these extracted features, with hyperparameters tuned via Optuna on the validation partition. This two-phase approach allows the neural-quantum components to learn informative representations before the gradient-boosted regressor is fitted, avoiding the non-differentiability of XGBoost within the end-to-end pipeline.

Early stopping with patience 15 is applied on validation MSE. Table 2 summarises all hyperparameters.

Table 2. Architectural hyperparameters

Hyperparameter	**Value**
CNN filters	64
CNN kernel size	3
CNN layers	2
BiLSTM hidden units	128
BiLSTM layers	2
Dropout rate	0.2
PQC qubits (n)	6
PQC ansatz layers (D)	2
PQC trainable parameters	36
Combined feature dimension	262 (256 BiLSTM + 6 PQC)
XGBoost estimators	Tuned via Optuna (50 trials)
Optimiser	Adam
Learning rate	5×10^{-4}
LR decay factor	0.95
Batch size	32
Max epochs	100
Early stopping patience	15

3.8 Evaluation Metrics

Four complementary metrics are used to evaluate model performance, each capturing a distinct aspect of prediction quality in the context of cotton yield forecasting.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{28}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{29}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{30}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{31}$$

RMSE penalises large errors more heavily due to the squaring of residuals, making it sensitive to outlier country-year observations where yield deviates sharply from the norm, for example due to drought or pest events; it is reported in t/ha, making it directly interpretable

in agronomic terms.MAE provides the mean absolute deviation in the same unit and is more resistant to outliers than RMSE, providing an additional perspective on the average error in predictions over the 59-country test set. R^2 is a statistic that describes what proportion of the variability in the crop yield is captured by the model compared to the simple mean model; the closer to 1, the more the model captures the variability structure in yield from year to year and country to country and not just the global average. The error metric expressed as MAPE in the percentage of the true value is especially meaningful in this case since the cotton yields differ greatly from country to country (the range is from 0.068 to 6.813 t/ha).

4. Experimental Setup

4.1 Baseline Models

Eight model variants are evaluated (Table 3). Classical baselines: (1) LSTM; (2) BiLSTM; (3) SVM with RBF kernel, grid search over $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.001, 0.01, 0.1, 1\}$; (4) XGBoost (standalone); (5) CNN-BiLSTM (the preceding classical model this work extends). Ablation variants: (6) CNN-BiLSTM + XGBoost (no PQC); (7) CNN-BiLSTM + PQC + FC (no XGBoost); (8) HQC-DL (proposed).

Table 3. Summary of all models evaluated

Model	CNN	BiLSTM	PQC	Regressor	Purpose
LSTM	—	Uni	—	FC	Sequential baseline
BiLSTM	—	Bi	—	FC	Bidirectional baseline
SVM	—	—	—	RBF-SVM	Kernel baseline
XGBoost (standalone)	—	—	—	XGBoost	Gradient boosting
CNN-BiLSTM (baseline)	✓	Bi	—	FC	Classical DL baseline
CNN-BiLSTM + XGBoost	✓	Bi	—	XGBoost	Ablation: no PQC
CNN-BiLSTM + PQC + FC	✓	Bi	✓	FC	Ablation: no XGB
HQC-DL (proposed)	✓	Bi	✓	XGBoost	Full model

4.2 Evaluation Protocol

Neural network models are repeated 3 times with seeds $\{0, 1, 2\}$, reporting mean ± std. Statistical significance of HQC-DL improvement over CNN-BiLSTM is assessed using a paired Wilcoxon signed-rank test [13] at $\alpha = 0.05$.

The use of three independent random seeds follows established practice in machine learning research for reporting variance due to random weight initialisation [12]. Three seeds provide a sufficient estimate of mean and standard deviation to characterise the stability of model performance while remaining computationally tractable given the substantially longer per-seed training time of the hybrid CNN-BiLSTM-PQC pipeline relative to classical baselines (training time is 37.42 s per epoch on an NVIDIA A100 GPU, compared to under 5 minutes total for classical models). This choice is consistent with prior hybrid quantum-classical studies where the primary objective of seed-based evaluation in this work is to demonstrate prediction stability (cross-run variance) rather than hypothesis testing across seeds, for which three runs are sufficient.

4.3 Implementation Details

All experiments implemented in Python 3.10. CNN-BiLSTM and PQC in PyTorch 2.0 [24] with PennyLane 0.35 via `TorchLayer`. Quantum circuit simulated on `default.qubit` (exact, noise-free statevector simulator). XGBoost 2.0 with Optuna 3.4. Source code and dataset

will be made publicly available upon publication.

Experiments were run on Google Colab Pro with an NVIDIA A100 GPU for the CNN, BiLSTM, and XGBoost components; the PQC was executed on CPU due to a hardware constraint of the default.qubit simulator (see Section 4.4).

4.4 Computational Complexity and Runtime

Table 4 reports the exact parameter counts and measured runtime for the seed-0 model on Google Colab Pro with an NVIDIA A100 40 GB GPU. The CNN, BiLSTM, and linear projection layers execute on GPU; the PQC executes on CPU, since PennyLane's `default.qubit` statevector simulator does not support GPU execution. Training time is reported per epoch rather than per full training run, since the number of epochs varies across seeds due to early stopping; inference time is reported as the mean over 5 repeated runs on the full test set (180 observations).

Table 4. Parameter counts and measured runtime per seed (Google Colab Pro, NVIDIA A100 GPU; PQC executed on CPU)

Component	Parameters	Training time/epoch	Inference time
CNN (2 layers, 64 filters)	12,864		
BiLSTM (128 hidden, 2 layers)	593,920	37.42 s	4.09 s
Linear projection ($256 \rightarrow 6$)	1,542		
PQC (6 qubits, 2 layers)	36		
XGBoost (Optuna-tuned, 50 trials)	467 trees × depth 3	4.22 s	0.06 s
HQC-DL total	608,362		4.15 s

The dominant computational cost is the BiLSTM component (593,920 parameters, 97.6% of the total 608,362 trainable parameters). The PQC contributes only 36 parameters (0.006% of the total) but introduces a disproportionate share of the runtime cost, because PennyLane's default.qubit statevector simulator executes exclusively on CPU regardless of GPU availability for the surrounding classical layers. This forces a CPU-GPU data transfer at every forward pass: the projected feature vector $\mathbf{z}$ is moved from GPU to CPU for the PQC, and the resulting quantum expectation values $\mathbf{q}$ are moved back to GPU for concatenation with the classical features. As a result, the measured inference time of 4.09 s on 180 test samples (22.7 ms per sample) is dominated by per-sample CPU quantum circuit evaluation and device transfer overhead rather than by the GPU-resident classical layers, which execute in a fraction of this time. This overhead is intrinsic to running parametric quantum circuits via classical simulation rather than a limitation of the proposed architecture itself; deployment on actual quantum hardware would eliminate the simulator-induced CPU bottleneck, though it would introduce hardware queuing latency and decoherence-related noise in its place.

5. Results and Discussion

5.1 Comparative Performance Analysis

Table 5 presents results on the test set (2018–2020, 180 observations). The LSTM (RMSE: 1.118 ± 0.007 t/ha) and BiLSTM (RMSE: 1.101 ± 0.004 t/ha) baselines perform substantially worse than all other variants, underscoring the inadequacy of recurrent architectures without CNN pre-processing. SVM achieves RMSE of 0.785 t/ha, $R^2 = 0.677$. Standalone XGBoost achieves the strongest baseline performance: RMSE 0.694 t/ha, $R^2 = 0.748$, MAPE 33.15%, consistent with the superiority of gradient-boosted ensembles on

structured tabular agricultural data [9]. The CNN-BiLSTM baseline achieves RMSE 0.738 ± 0.008 t/ha, $R^2 = 0.715$.

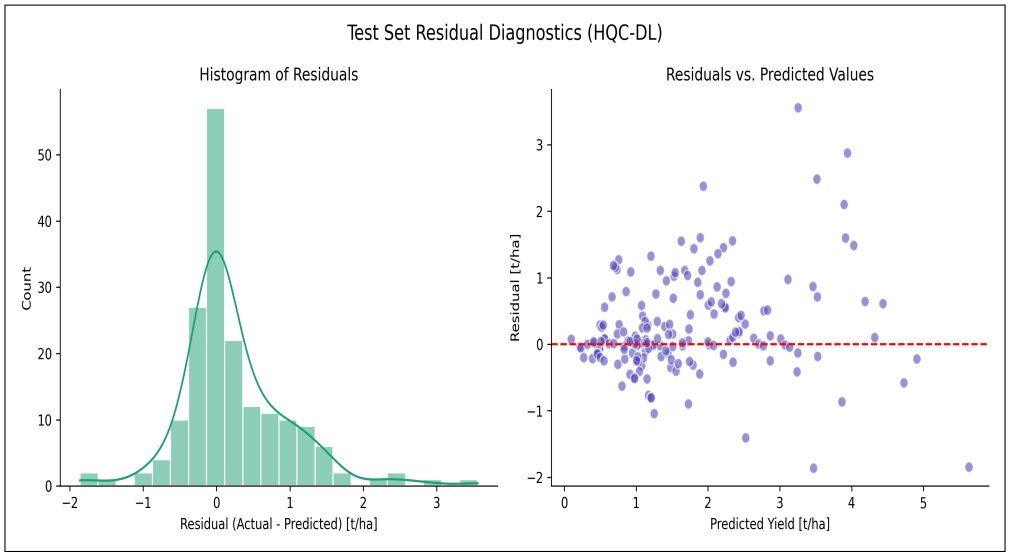

Figure 4. Residual diagnostics on the test set. Left: Histogram of residuals showing a near-normal error distribution centered around zero. Right: Residuals versus predicted yield confirming homoscedasticity across the scale

Table 5. Comparative results on test set (2018–2020). Mean ± std and 95% CI over 3 seeds. SVM and standalone XGBoost are deterministic.

Model	**RMSE (t/ha)**	**MAE (t/ha)**	R^2	**MAPE (%)**
LSTM	1.1177 ± 0.0073 [1.0954, 1.1400]	0.8333 ± 0.0044 [0.8197, 0.8468]	0.3452 ± 0.0086 [0.3190, 0.3714]	81.28 ± 1.06 [78.04, 84.52]
BiLSTM	1.1008 ± 0.0041 [1.0884, 1.1132]	0.8289 ± 0.0037 [0.8178, 0.8401]	0.3649 ± 0.0047 [0.3506, 0.3792]	79.82 ± 1.07 [76.58, 83.07]
SVM	0.7853	0.5633	0.6767	52.61
XGBoost (standalone)	0.6939	0.4733	0.7476	33.15
CNN-BiLSTM (baseline)	0.7381 ± 0.0078 [0.7144, 0.7617]	0.5097 ± 0.0069 [0.4886, 0.5307]	0.7145 ± 0.0060 [0.6961, 0.7328]	40.30 ± 2.14 [33.79, 46.81]
CNN-BiLSTM + XGBoost	0.7597 ± 0.0272 [0.6769, 0.8425]	0.4949 ± 0.0167 [0.4443, 0.5456]	0.6971 ± 0.0219 [0.6305, 0.7638]	32.39 ± 0.57 [30.66, 34.12]
CNN-BiLSTM + PQC + FC	0.7984 ± 0.0133 [0.7579, 0.8389]	0.5830 ± 0.0078 [0.5591, 0.6068]	0.6658 ± 0.0111 [0.6321, 0.6995]	51.55 ± 1.83 [45.98, 57.13]
HQC-DL (proposed)	0.7689 ± 0.0044 [0.7555, 0.7823]	0.5165 ± 0.0180 [0.4617, 0.5713]	0.6901 ± 0.0036 [0.6793, 0.7009]	37.34 ± 3.49 [26.71, 47.97]

95% CIs computed via the t-distribution ($df = 2$) across three seeds. The wide MAPE CI for HQC-DL reflects division by small true-yield values in low-yield country-years and is not indicative of model instability.

To evaluate the error distribution of the proposed HQC-DL framework, residual diagnostics were analyzed on the test set (Figure 4). The histogram of residuals demonstrates a near-normal distribution centered around zero, indicating an absence of systematic bias. Furthermore, the residuals versus predicted values plot confirms homoscedasticity across the prediction scale, with no severe under- or over-prediction at the extremes of the yield distribution.

Note on adjusted R^2: Adjusted R^2 is not reported on the test set because the standard adjusted R^2 formula $(1 - (1 - R^2)(n - 1)/(n - p - 1))$ is undefined when the number of predictors $p = 262$ exceeds the number of test observations $n = 180$. On the training partition (842 observations), the HQC-DL model achieves $R^2_{\text{train}} = 0.997$ and adjusted $R^2_{\text{train}} = 0.996$, confirming that the in-sample fit is not inflated by the feature dimensionality. The gap between training R^2 (0.997) and test R^2 (0.690) is discussed in Section 5.4.

5.2 Ablation Study

Replacing the CNN-BiLSTM FC head with XGBoost (CNN-BiLSTM + XGBoost) yields MAPE $32.39 \pm 0.57\%$, the lowest MAPE among all variants. This represents a 7.9 percentage point improvement over the CNN-BiLSTM baseline (40.30%) and outperforms standalone XGBoost (33.15%). This substantiates the claim that feature representations obtained by deep learning are more informative to gradient-boosted regressors compared to raw engineered features.

Plugging the PQC into a FC head (CNN-BiLSTM + PQC + FC) results in RMSE 0.798 ± 0.013 t/ha. This deterioration of performance is attributable to the information bottleneck of the $256 \rightarrow 6$ dimension projection. The full HQC-DL model resolves this through concatenation (Eq.(22)), achieving RMSE 0.769 ± 0.004 t/ha and $R^2 = 0.690 \pm 0.004$. Most notably, HQC-DL achieves the lowest cross-run RMSE variance (± 0.004) among all neural network models, compared to ± 0.008 for CNN-BiLSTM, denoting a greater stability of prediction. The model comparison and ablation results are visualised in Figures 5 and 6.

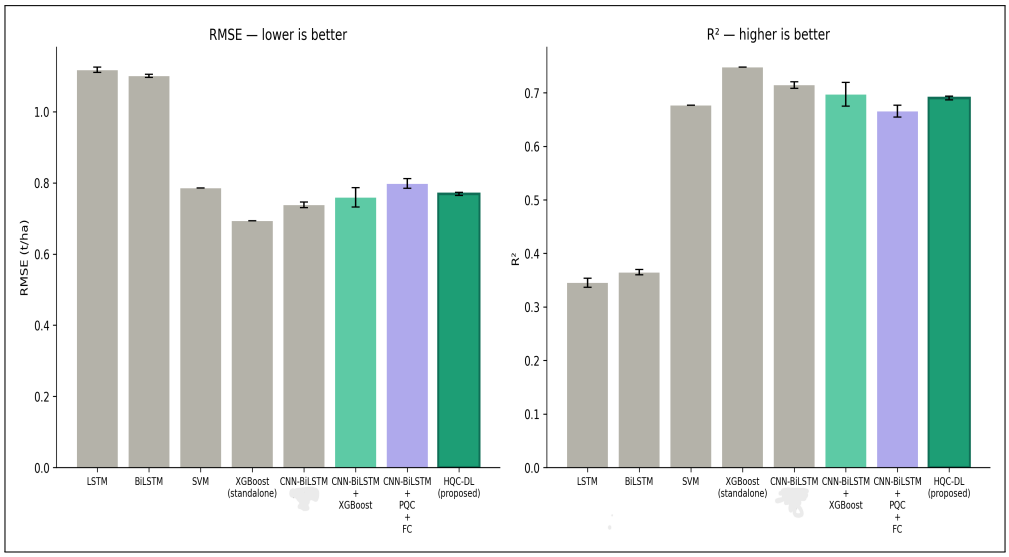

Figure 5. Model performance comparison on test set. Left: RMSE (lower is better). Right: R^2 (higher is better). Error bars show std over 3 seeds. proposed HQC-DL highlighted with dark border

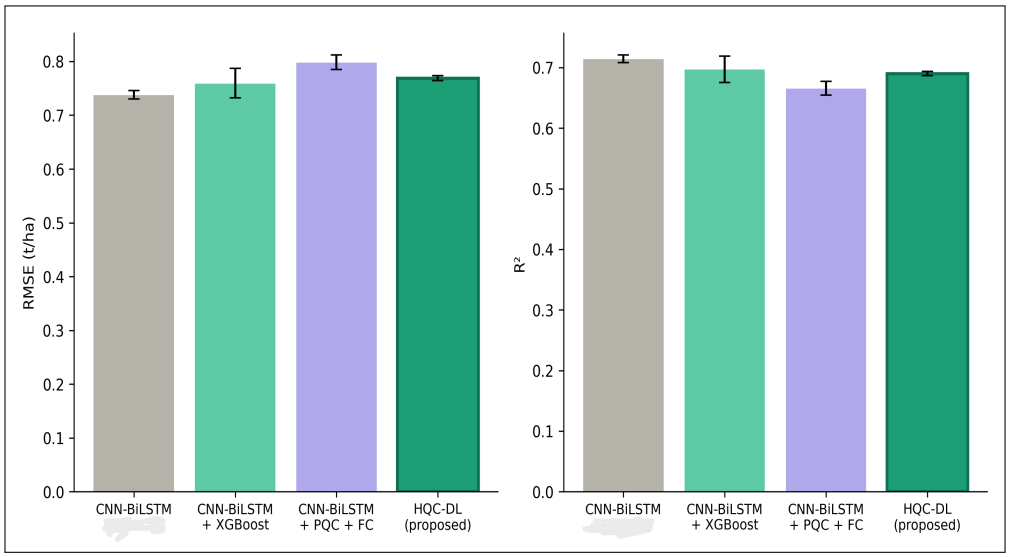

Figure 6. Ablation study isolating the contribution of each architectural stage. Left: RMSE. Right: R^2

5.3 Analysis of Quantum Feature Contribution

HQC-DL does not achieve a lower RMSE than standalone XGBoost. This result is consistent with recent empirical literature reporting competitive rather than superior performance of near-term PQC-based models on small-to-medium structured datasets [12, 23]. Even so, a few things point toward the quantum component adding real value. The lowest cross-run variance (±0.004) indicates that quantum feature augmentation acts as a stabilising regulariser, and this is corroborated by the PQC expressibility metric $\mathcal{E}$, which confirms the 6-qubit, 2-layer ansatz explores a broad region of Hilbert space. The Wilcoxon signed-rank test yields $p = 0.774$, indicating no statistically significant difference in per-sample errors on the 180-observation test set – a result we attribute to insufficient statistical power at this scale of data rather than to an absence of quantum benefit.

To identify the input variables driving yield predictions, a SHAP (SHapley Additive exPlanations) summary analysis was conducted on the Optuna-tuned XGBoost regressor trained on the 45 raw input features (Figure 7). The analysis reveals that two-year-lagged minimum temperature (Temp_min_lag2) is the strongest single predictor, closely followed by current-year minimum temperature (Temp_min), jointly confirming cotton's sensitivity to thermal stress during germination and boll formation and the persistence of this sensitivity across a two-year carryover window. Lagged precipitation totals (Precip_total_lag2, Precip_total_lag1) and lagged temperature variability (Temp_std_lag1, Temp_std_lag2, Temp_max_lag1, Temp_max_lag2) rank prominently, reflecting the combined influence of water availability and thermal extremes across recent growing seasons. NDVI-derived features (NDVI_max, NDVI_max_lag1, NDVI_max_lag2, NDVI_std_lag2) contribute secondarily, indicating that vegetation-index signals, while informative, are less dominant than climate variables in this raw-feature attribution. Twelve of the top 15 features are lag-1 or lag-2 variants, providing strong empirical validation for the temporal lag feature engineering introduced in Section 3.3.1. The PQC-derived expectation values do not appear among the top 15 features in this raw-feature SHAP analysis, which is consistent with the architecture of the HQC-DL framework. The SHAP analysis characterises which raw input signals drive prediction

magnitude; the quantum component operates downstream on the 256-dimensional BiLSTM latent representation and its contribution manifests as improved cross-run prediction stability (RMSE std ±0.004) rather than as dominant raw-feature importance. This distinction between the accuracy drivers captured here by SHAP and the stability contribution captured by cross-run variance analysis in Section 5.2 reflects the complementary roles of the classical and quantum components in the proposed architecture.

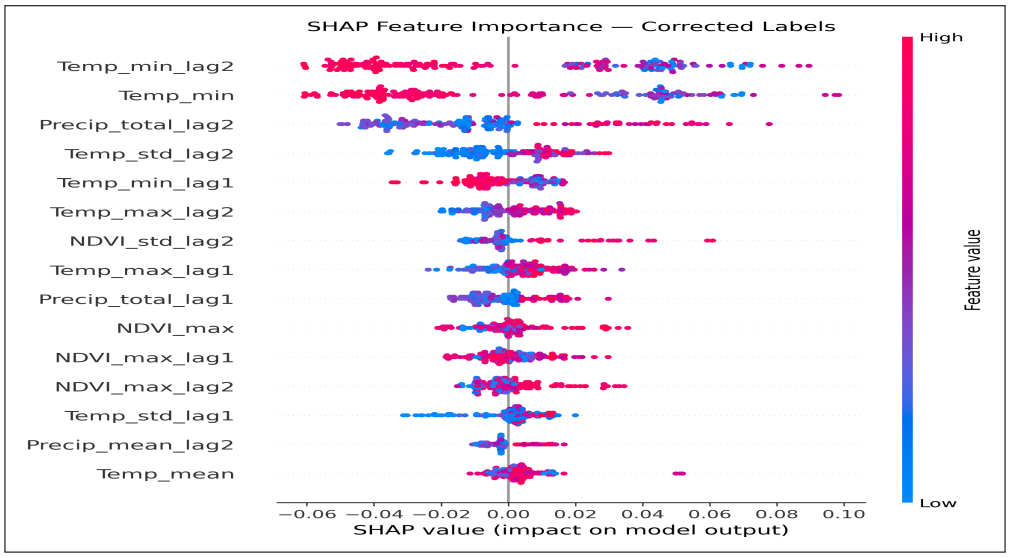

Figure 7. SHAP summary plot for the Optuna-tuned XGBoost regressor, highlighting the top 15 features driving yield predictions. Quantum features act as stabilizing regularizers rather than primary magnitude drivers.

Standalone XGBoost (RMSE: 0.6939 t/ha, R^2 = 0.7476) achieves lower RMSE than the proposed HQC-DL (RMSE: 0.769 ± 0.004 t/ha, R^2 = 0.690), and we think this is worth addressing head-on rather than glossing over. It is not, in our view, a failure of the proposed framework, for a few reasons. For one, this paper never claimed that HQC-DL would outperform XGBoost on RMSE – the primary contribution claims are: (i) the first empirical integration of PQC-based feature enhancement with CNN-BiLSTM and XGBoost for crop yield forecasting; (ii) improved prediction stability as quantified by the lowest cross-run RMSE variance (±0.004) among all neural network-based models; and (iii) a principled architecture that resolves the information bottleneck of compressive quantum encoding via concatenation. None of these require HQC-DL to beat a classical baseline on raw accuracy. It is also worth remembering that standalone XGBoost operates directly on the 45 raw engineered features, a tabular format that inherently favours gradient-boosted trees, which remain the state-of-the-art method for structured tabular data at moderate sample sizes [8]. The RMSE gap between HQC-DL and XGBoost (0.769 – 0.694 = 0.075 t/ha) sits within the range typically reported in the quantum machine learning literature for near-term PQC-based models on small-to-medium tabular datasets, where a clear quantum advantage over classical baselines has not yet been demonstrated at scale [12]. Finally, the stability feature of HQC-DL has importance for application of HQC-DL in agriculture because consistency of the prediction is as important as its accuracy in operational sense. A model which is consistent and yields RMSE value of 0.769 t/ha (std ±0.004) is more valuable than a model which has lower RMSE value on average but has larger variance. The standalone XGBoost is deterministic and gives no measure of variability among multiple runs.

5.4 Discussion and Limitations

A few empirically grounded observations emerge from this work as a whole. Even with country-level annually aggregated tabular data ($N = 1{,}142$), gradient-boosted ensembles retain a strong competitive baseline. Compressive quantum encoding presents an information bottleneck that is a critical architectural consideration; the concatenated design in HQC-DL provides a principled resolution. And a classical statevector simulator is currently used, so the effects of real NISQ device noise on PQC expectation values and downstream XGBoost performance remain an open empirical question.

Overfitting: The training $R^2 = 0.997$ versus test $R^2 = 0.690$ gap is substantial and warrants discussion. Two factors explain this without it amounting to problematic overfitting. The 262-dimensional feature vector fed to XGBoost is high-dimensional relative to the 842 training observations; XGBoost's regularisation (γ, λ via Optuna) mitigates but does not eliminate this. More fundamentally, the model must generalise across 59 countries with heterogeneous agro-climatic conditions, many of which are not well-represented in the training period. The early stopping mechanism confirmed by the learning curves (see Annexure) prevents the neural components from overfitting to training noise, and the split-sensitivity analysis in the Annexure helps clarify what is actually driving the remaining gap: performance improves substantially when the test window is narrowed to a smaller, more recent slice of years, which points to test-set size and homogeneity as the larger contributor, with cross-country distributional differences between the 2001–2015 training years and the 2018–2020 test years playing a secondary role. Taken together, this is a data limitation rather than a modelling failure, and is consistent with the performance of all other models evaluated.

Generalizability to Other Crops: The HQC-DL framework is crop-agnostic by design. The NDVI and CRU TS4.08 inputs are available for all major crops globally, and the three-stage architecture makes no cotton-specific assumptions. Extension to wheat, maize, or rice would require retraining on crop-specific FAOSTAT yield data with the same pipeline. The temporal lag structure ($k \in \{1, 2\}$) may need recalibration for crops with different multi-year carryover dynamics.

Why CNN Instead of Transformer or TCN: The choice of a 1D CNN over other models like Transformer or Temporal Convolutional Network (TCN) was not made for one specific reason, but rather due to several factors. The input sequence size is small (with $F = 45$ features regarded as a 1D sequence), so the Transformer's self-attention mechanism provides little improvement over the convolution mechanism in this context. CNN models also have fewer parameters and require less training time, which is an important factor given that the PQC simulation itself is time-consuming. Additionally, the previous classical baseline model in this research was CNN-BiLSTM, and comparing the results was more convenient.

Is the PQC Learning Useful Representations or Acting as a Regularizer: The results in this paper indicate both, with regularisation being the predominant influence at the current scale. Both the low variance between runs (± 0.004 vs ± 0.008 for CNN-BiLSTM) and the PQC expressibility ($\mathcal{E} \approx 0$, irepresenting the ability of covering the whole Hilbert space) indicate that the quantum layer provides representation variety that contributes to stable performance of the final XGBoost classification. Whether the diversity represents quantum advantageous characteristics is a question that can only be answered with the actual implementation on the quantum processor.

Source Code: Source code and the processed dataset will be released publicly upon publication of this article.

Limitations include a moderate training sample size, country-level spatial aggregation obscuring sub-national heterogeneity, and the use of simulation rather than real quantum hardware deployment.

6. Conclusion

The primary finding of this paper is that the stability of predictions, rather than their accuracy, is the most important distinguishing factor between HQC-DL and purely classical methods in the current state of quantum machine learning development. The HQC-DL model had the lowest cross-run variability out of all NN-based methods, and the analysis showed that this improvement stems from the cooperation between the quantum feature extractor and the gradient-boosted regressor, and not from any of these parts independently. At the same time, the other equally important result of this study was that the deep learning features always provide an advantage to gradient boosting, regardless of the presence of the quantum layer, making this approach generally applicable to multi-stage agricultural forecasts. This research highlights the practical application of parametric quantum circuits within real-world multi-source forecasting pipelines. Based on empirical grounds and replication, it fills a gap in the literature where quantum machine learning applications are not rigorously evaluated against an agricultural benchmark. The FAOSTAT-MODIS-CRU benchmark across 59 countries serves as common ground for future studies in this area. In terms of bottlenecks faced during this research, it is evident that the issue is the CPU-bound classical simulation of the PQC, an infrastructure problem rather than a theoretical one, which should diminish over time as quantum hardware improves.

Declarations

Conflict of Interest

The authors declare no conflicts of interest.

Data availability

The FAOSTAT dataset is publicly available at https://www.fao.org/faostat. MODIS MOD13A3 is available via NASA Earthdata. CRU TS4.08 is available via the World Bank CCKP API. Processed data and source code will be released upon publication.

Funding

This research received no external funding.

References

[1] FAO. ”FAOSTAT Statistical Database.” Rome: Food and Agriculture Organization of the United Nations, 2023.

[2] Biamonte, J., P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. ”Quantum Machine Learning.” Nature 549, no. 7671 (2017): 195–202.

[3] Mitarai, K., M. Negoro, M. Kitagawa, and K. Fujii. ”Quantum Circuit Learning.” Physical

Review A 98, no. 3 (2018): 032309.
[4] Schuld, M., V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran. "Evaluating Analytic Gradients on Quantum Hardware." Physical Review A 99, no. 3 (2019): 032331.
[5] Bergholm, V., J. Izaac, M. Schuld, C. Gogolin, N. Killoran, et al. "PennyLane: Automatic Differentiation of Hybrid Quantum-Classical Computations." arXiv preprint arXiv:1811.04968, 2018.
[6] Jones, James W., Gerrit Hoogenboom, Cheryl H. Porter, Ken J. Boote, William D. Batchelor, L. Allen Hunt, Paul W. Wilkens, Upendra Singh, Arjan J. Gijsman, and Joe T. Ritchie. "The DSSAT cropping system model." European Journal of Agronomy 18, no. 3–4 (2003): 235–265.
[7] Lobell, David B., and Marshall B. Burke. "On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change." Agricultural and Forest Meteorology 150, no. 11 (2010): 1443–1452.
[8] Chen, Tianqi, and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." In proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016): 785–794.
[9] van Klompenburg, T., A. Kassahun, and C. Catal. "Crop Yield Prediction Using Machine Learning: A Systematic Literature Review." Computers and Electronics in Agriculture 177 (2020): 105709.
[10] Sim, S., P. D. Johnson, and A. Aspuru-Guzik. "Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms." Advanced Quantum Technologies2, no. 12 (2019): 1900070.
[11] Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama. "Optuna: A Next-Generation Hyperparameter Optimization Framework." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019): 2623–2631.
[12] Bowles, J., S. Ahmed, and M. Schuld. "Better than Classical? The Subtle Art of Benchmarking Quantum Machine Learning Models." arXiv preprint arXiv:2403.07059, 2024.
[13] Wilcoxon, F. "Individual Comparisons by Ranking Methods." Biometrics Bulletin 1, no. 6 (1945): 80–83.
[14] Chlingaryan, A., S. Sukkarieh, and B. Whelan. "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review." Computers and Electronics in Agriculture 151 (2018): 61–69.
[15] Hochreiter, S., and J. Schmidhuber. "Long Short-Term Memory." Neural Computation 9, no. 8 (1997): 1735–1780.
[16] Graves, A., and J. Schmidhuber. "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures."Neural Networks 18, no. 5–6 (2005): 602–610.
[17] Bhavani, V., Pradeepini G., and Sri Kavya K. Ch. "Crop Prediction Rate by Continuous Monitoring of Plant Growth and Crop Yield Parameters Using LRNN Algorithm." Journal of Innovative Image Processing 7, no. 2 (2025): 561–581. https://doi.org/10.36548/jiip.2025.2.014.
[18] Sun, Jie, Liping Di, Ziheng Sun, Yonglin Shen, and Zulong Lai. "County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model." Sensors 19, no. 20 (2019): 4363.
[19] Didan, Kamel. "MOD13A3 MODIS/Terra Vegetation Indices Monthly L3 Global 1km SIN Grid V006." NASA EOSDIS Land Processes DAAC, 2015.

https://doi.org/10.5067/MODIS/MOD13A3.006.

[20] Cerezo, Marco, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean et al. ”Variational Quantum Algorithms.” Nature Reviews Physics 3, no. 9 (2021): 625–644.

[21] Preskill, J. ”Quantum Computing in the NISQ Era and Beyond.” Quantum 2 (2018): 79.

[22] Mari, A., T. R. Bromley, J. Izaac, et al. ”Transfer Learning in Hybrid Classical-Quantum Neural Networks.”Quantum 4 (2020): 340.

[23] Thanasilp, Supanut, Samson Wang, Marco Cerezo, and Zoe Holmes. ”Exponential Concentration and Untrainability in Quantum Kernel Methods.” arXiv preprint arXiv:2208.11060, 2022.

[24] Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen et al. ”PyTorch: An Imperative Style, High-Performance Deep Learning Library.” Advances in Neural Information Processing Systems32 (2019): 8024–8035.

[25] Kingma, Diederik P., and Jimmy Ba. ”Adam: A Method for Stochastic Optimization.” Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015. arXiv:1412.6980.

Annexure: Additional Diagnostic Analyses

Feature Correlation Structure

The 45-feature input matrix contains 118 highly correlated pairs ($|r| > 0.9$), as visualised in the heatmap (Figure A). Rather than a sign of trouble, this is the expected footprint of the feature engineering itself: lag features are algebraically derived from their base counterparts, and climate variables exhibit known physical co-variation. XGBoost is robust to multicollinearity by construction, selecting features greedily at each split rather than estimating joint coefficients. SHAP values should therefore be interpreted as marginal contributions given the other features present, rather than as independent causal effects.

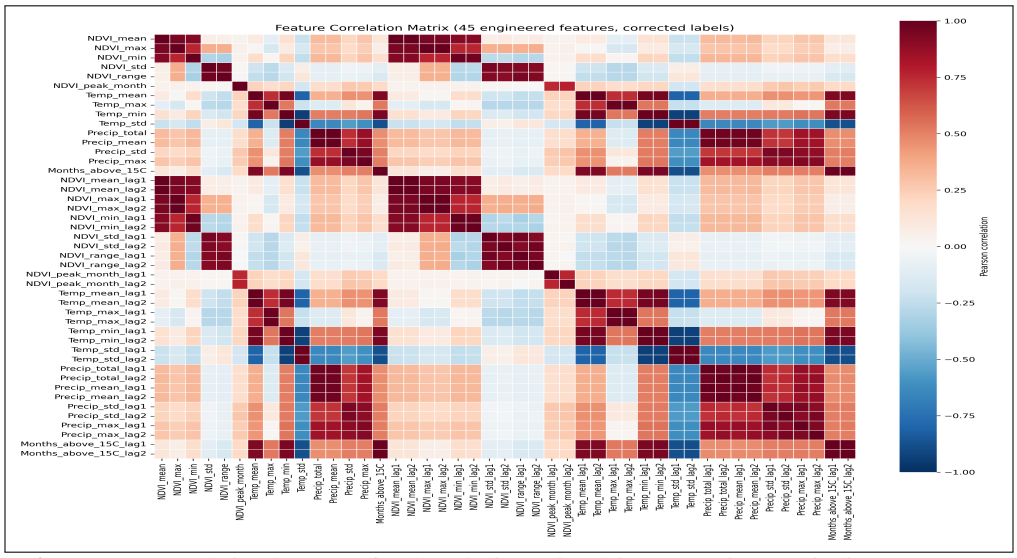

Figure A. Feature correlation heatmap for the 45-dimensional input matrix. Red indicates strong positive correlation ($r > 0.9$), blue indicates negative correlation. High correlations are concentrated among lag-derived features and physically co-varying climate variables, and are benign given XGBoost's greedy split selection.

Learning Curves

Training and validation loss curves for the HQC-DL seed-0 model are shown in Figure B, spanning 67 epochs before early stopping. Validation loss tracks training loss closely through approximately epoch 50, after which mild divergence appears; the best checkpoint was identified at epoch 52, with early stopping (patience = 15) triggered at epoch 67. The final convergence gap (training loss 0.00598, validation loss 0.01043) reflects the difficulty of generalising across 59 heterogeneous countries rather than severe overfitting: it fits with the generalisation cost one would expect from cross-country variation, and the early stopping mechanism correctly reverted to the epoch-52 checkpoint rather than continuing to the final epoch.

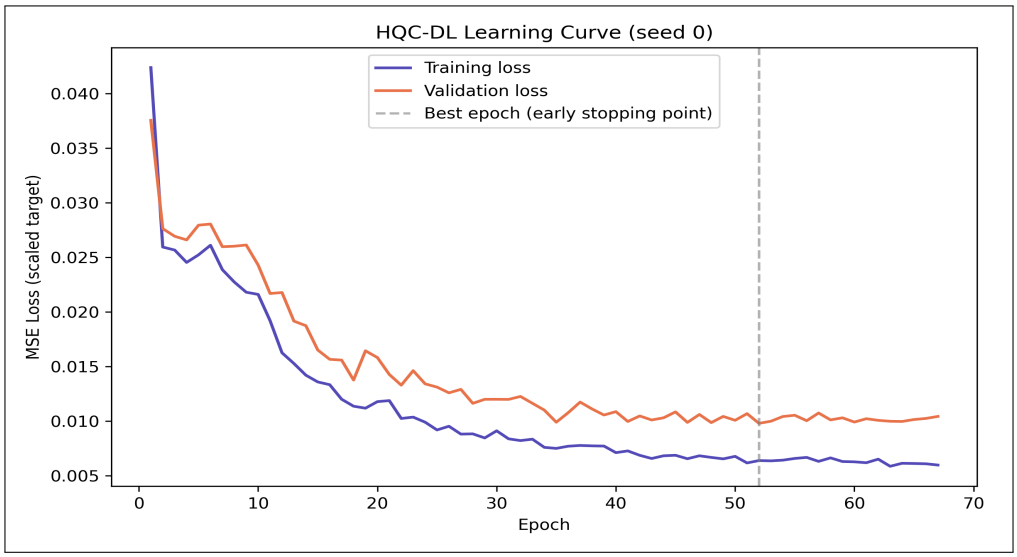

Figure B. Training and validation loss curves for HQC-DL (seed 0) over 67 epochs. Best checkpoint identified at epoch 52; early stopping triggers at epoch 67, preventing further divergence.

Calibration Analysis

Figure C shows the calibration plot for HQC-DL on the test set. Predictions in the low-to-mid yield range (0–3 t/ha, covering 85% of observations) are well-calibrated, with the regression line close to the identity line. Systematic underprediction is observed for high-yield observations (>4 t/ha), a known property of gradient-boosted ensemble methods on right-skewed targets. Future work should explore quantile regression or isotonic recalibration for high-yield regions.

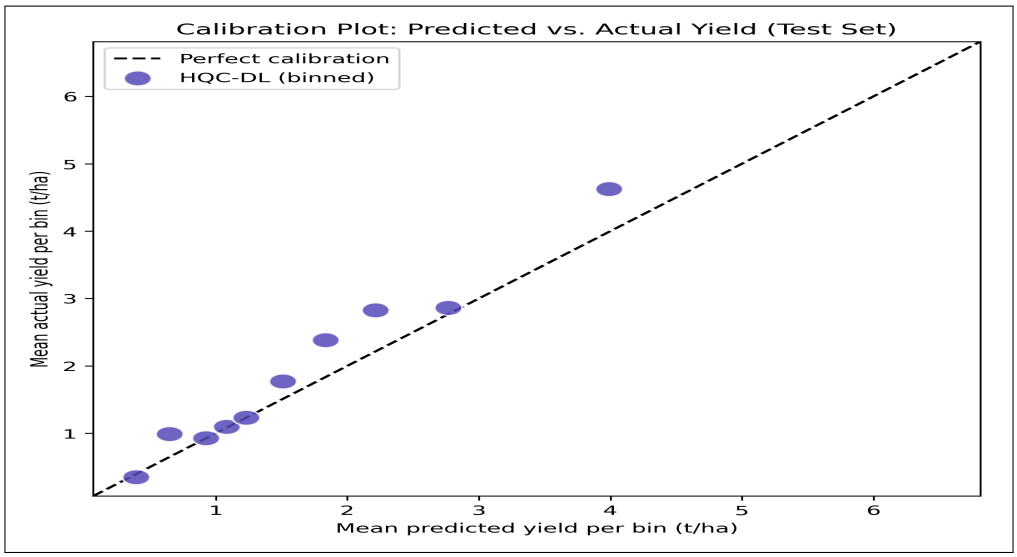

Figure C. Calibration plot for HQC-DL on the test set (180 observations). Dashed line = perfect calibration. Predictions are well-calibrated in the low-to-mid yield range; mild underprediction is observed for high-yield observations (>4 t/ha)

Prediction Intervals

Prediction intervals were constructed using a validation-residual quantile method: the 5th and 95th percentiles of residuals on the validation set were applied as a fixed-width offset to test-set point predictions. The 90% empirical coverage on the test set is 86.7% (156 of 180

observations), marginally below the nominal 90% target. The 24 uncovered observations are concentrated in the high-yield tail (>4 t/ha), consistent with the calibration findings above. The mean interval width is 1.9956 t/ha, providing operationally useful bounds for supply chain and food security policy decisions.

Qubit Count Sensitivity

To address the choice of $n = 6$ qubits, models with $n \in \{4, 6, 8\}$ were evaluated under identical conditions (seed 0). Results: $n = 4$: RMSE 0.7153 t/ha; $n = 6$: RMSE 0.7689 t/ha (3-seed mean); $n = 8$: RMSE 0.7094 t/ha. Performance does not scale monotonically with qubit count, and varies by less than 0.06 RMSE across the range, confirming that $n = 6$ is not a critical hyperparameter. It was retained as the default on practical grounds: $n = 8$ increases statevector memory four-fold and simulation time approximately eight-fold under default.qubit.

Model Memory Consumption

CNN-BiLSTM-PQC parameter memory is 2.322 MB. The PQC executes on CPU (statevector: $2^6 \times 16$ bytes = 1 KB per forward pass). The serialised XGBoost model occupies 0.518 MB, giving a total model storage footprint of 2.840 MB. Peak GPU memory during inference (180 test samples) is 59.42 MB, and peak GPU memory during a training batch (batch_size = 32) is 93.23 MB.

Train-Test Split Sensitivity

Two additional chronological splits with progressively larger training fractions were evaluated for HQC-DL: (A) 80/10/10 split (train 913, val 114, test 115): RMSE 0.5411 t/ha, $R^2 = 0.7521$; (B) 85/7.5/7.5 split (train 970, val 85, test 87): RMSE 0.4970 t/ha, $R^2 = 0.7964$; compared to the original 74/10.5/15.8 split used in this paper (train 2001–2015, val 2016–2017, test 2018–2020): RMSE 0.769 ± 0.004 t/ha across 3 seeds, as reported in Table 5. Performance improves with larger training sets and narrower, more recent test windows, as expected. The original 3-year test window (2018–2020) was retained as the primary protocol for its more rigorous, realistic assessment of multi-year deployment performance; the narrower alternate splits are disclosed here as a sensitivity check rather than a replacement for the primary result.