APPENDIX

1 LSTM AND QLSTM DETAILS

This appendix provides details on the LSTM and QLSTM model architectures used in the study.

1.1 LSTM

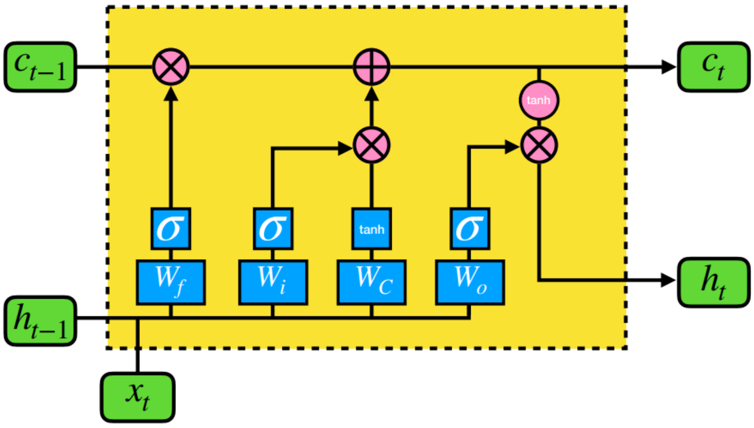

Figure 1. LSTM Circuit (11)

The LSTM architecture used in this study stacks multiple LSTM cells to model long-term dependencies. The information flow in an LSTM cell is described by the equations:

$$
\begin{aligned}
f_t &= \sigma(W_f \cdot v_t + b_f), \\
i_t &= \sigma(W_i \cdot v_t + b_i), \\
C_t &= \tanh(W_C \cdot v_t + b_C), \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot C_t, \\
h_t &= o_t \cdot \tanh(c_t),
\end{aligned}
$$

where σ denotes the sigmoid activation function, W and b are learnable parameters, f is the forget gate, i is the input gate, C is the cell state, c is the hidden state, and o is the output gate. The LSTM was chosen due to its proven ability to model sequence data across various domains. The LSTM cell architecture is illustrated as follows:

1.2 QLSTM

The QLSTM replaces LSTM cells with 6 variational quantum circuits (VQCs) to form a quantum LSTM cell. VQCs leverage a small number of qubits and gates to represent complex functions. This quantum layer showed quicker convergence and more stable loss than the classical LSTM (11). The information flow in a quantum LSTM cell is described by the equations:

$$
\begin{aligned}
f_t &= \sigma(\mathrm{VQC}_1(v_t)), \\
i_t &= \sigma(\mathrm{VQC}_2(v_t)), \\
C_t &= \tanh(\mathrm{VQC}_3(v_t)), \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot C_t, \\
h_t &= \mathrm{VQC}_5(o_t \cdot \tanh(c_t)).
\end{aligned}
$$

The VQC_x represent different quantum circuits used in the hybrid model. The QLSTM cell architecture is depicted as follows:

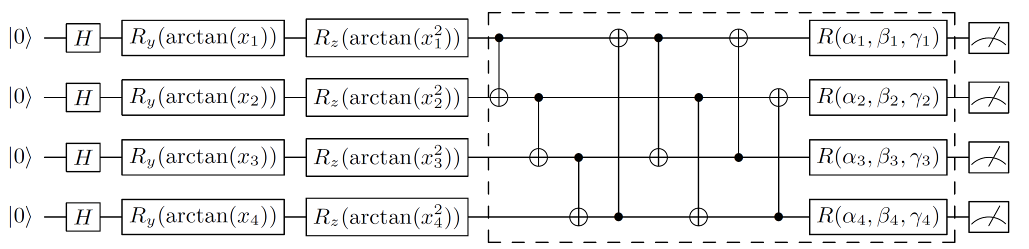

Figure 2. Generic VQC architecture for QLSTM. It consists of three layers: the data encoding layer (with the H, Ry, and Rz gates), the variational layer (dashed box), and the quantum measurement layer. (11)

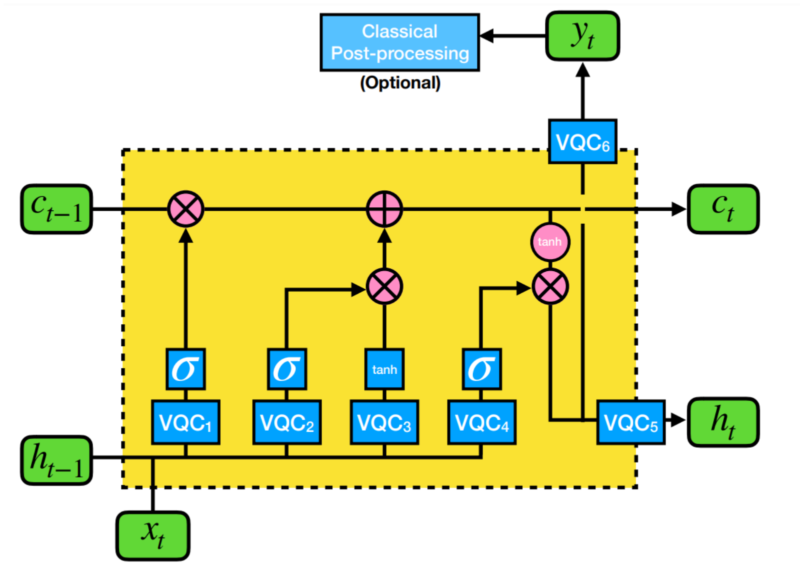

Figure 3. QLSTM Circuit (11)

2 EVALUATION METHODOLOGY

This appendix provides specifics on the quantitative metrics and procedures used to evaluate the LSTM and QLSTM model performance on the solar forecasting task.

2.1 Evaluation Metrics

The following quantitative metrics were computed to assess model accuracy:

- **Mean Absolute Error (MAE)**: Measures average absolute difference between predicted and actual values. Gives an indication of overall error. Lower is better.

$$\text{MAE} = \frac{1}{N}\sum |y_i - \hat{y}_i|$$

- **Mean Squared Error (MSE)**: Computes average squared difference between predicted and actual values. More sensitive to outliers than MAE. Lower is better.

$$\text{MSE} = \frac{1}{N}\sum (y_i - \hat{y}_i)^2$$

- **Root Mean Squared Error (RMSE)**: Square root of MSE. Allows interpretability in units of the target variable. Lower is better.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

- **T-statistic**: The T-statistic is a measure used to determine if there is a significant difference between the means of two groups. It is calculated as the difference between the sample means divided by the standard error of the difference between the means. The formula is given by:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{2}{n}}}$$

  Where $\bar{X}_1$ and $\bar{X}_2$ are the sample means, s_p is the pooled standard deviation, and n is the sample size for each group.
- **P-value**: The p-value is a fundamental concept in hypothesis testing. It represents the probability that the observed data (or something more extreme) would occur if the null hypothesis were true. A smaller p-value typically indicates stronger evidence against the null hypothesis. Conventionally, a p-value below 0.05 is considered statistically significant.
- **Effect Size (Cohen's *d*)**: While the T-statistic tells us if there is a statistically significant difference between groups, effect size quantifies the size of this difference. One commonly used measure is Cohen's d, calculated as:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

  Where s_p is the pooled standard deviation. Cohen's d values can be interpreted as small (0.2), medium (0.5), and large (0.8) effects.

These metrics were selected as standard measures of predictive accuracy for time series forecasting problems. MAPE was included due to its interpretability for solar power production. RMSE and R^2 were used as primary metrics for model comparison.

2.2 Evaluation Procedure

Metrics were computed on scaled predictions compared to scaled actual values for both the training and test sets. This enabled directly evaluating model generalization. Statistical significance testing using a paired t-test on RMSE values was also conducted to assess whether differences in LSTM and QLSTM errors were statistically significant. Model loss curves, prediction plots, and other visualizations were generated to provide qualitative evaluation.

By leveraging both quantitative metrics and qualitative assessments on scaled holdout data, this methodology enabled thoroughly evaluating how effectively the models learned to generalize. The comparative analysis focused on assessing whether the QLSTM architecture demonstrated significantly improved accuracy over classical LSTM for real-world solar forecasting.