QuLTSF: Long-Term Time Series Forecasting with Quantum Machine Learning

Hari Hara Suthan Chittoor[1][a], Paul Robert Griffin[1][b], Ariel Neufeld[2][c], Jayne Thompson[3,4][d] and Mile Gu[2,5][e]

[1]*School of Computing and Information Systems, Singapore Management University, 178902, Singapore*
[2]*Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*
[3]*Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore*
[4]
[5]*Centre for Quantum Technologies, National University of Singapore, Singapore*
{*haric, paulgriffin*}*@smu.edu.sg, ariel.neufeld@ntu.edu.sg, thompson.jayne2@gmail.com, mgu@quantumcomplexity.org*

Keywords: Quantum Computing, Machine Learning, Time Series Forecasting, Hybrid Model.

Abstract: Long-term time series forecasting (LTSF) involves predicting a large number of future values of a time series based on the past values. This is an essential task in a wide range of domains including weather forecasting, stock market analysis and disease outbreak prediction. Over the decades LTSF algorithms have transitioned from statistical models to deep learning models like transformer models. Despite the complex architecture of transformer based LTSF models 'Are Transformers Effective for Time Series Forecasting? (Zeng et al., 2023)' showed that simple linear models can outperform the state-of-the-art transformer based LTSF models. Recently, quantum machine learning (QML) is evolving as a domain to enhance the capabilities of classical machine learning models. In this paper we initiate the application of QML to LTSF problems by proposing *QuLTSF*, a simple hybrid QML model for multivariate LTSF. Through extensive experiments on a widely used weather dataset we show the advantages of QuLTSF over the state-of-the-art classical linear models, in terms of reduced mean squared error and mean absolute error.

1 INTRODUCTION

Time series forecasting (TSF) is the process of predicting future values of a variable using its historical data. TSF is an import problem in many fields like weather forecasting, finance, power management etc. There are broadly two approaches to handle TSF problems: statistical models and deep learning models. Statistical models, like ARIMA, are the traditional work horse for TSF since the 1970's (Hyndman, 2018; Hamilton, 2020). Deep learning models, like recurrent neural networks (RNN's), often outperform statistical models in large-scale datasets (Lim and Zohren, 2021).

Increasing the prediction horizon strain's the models predictive capacity. The prediction length of more than 48 future points is generally considered as *long-term time series forecasting* (LTSF) (Zhou et al., 2021). Transformers and attention mechanism proposed in (Vaswani, 2017) gained a lot of attraction to model sequence data like language, speech etc. There is a surge in the application of transformers to LTSF leading to several time series transformer models (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022; Liu et al., 2022; Wen et al., 2023). Despite the complicated design of transformer based models for LTSF problems, (Zeng et al., 2023) showed that a simple autoregressive model with a linear fully connected layer can outperform the state-of-the-art transformer models.

Quantum machine learning (QML) is an emerging field that combines quantum computing and machine learning to enhance tasks like classification, regression, generative modeling etc., using the currently available *noisy intermediate-scale quantum* (NISQ) computers (Preskill, 2018; Schuld and Petruccione, 2021; Simeone, 2022). Hybrid models contain-

[a] https://orcid.org/0000-0003-1363-606X
[b] https://orcid.org/0000-0003-2294-5980
[c] https://orcid.org/0000-0001-5500-5245
[d] https://orcid.org/0000-0002-3746-244X
[e] https://orcid.org/0000-0002-5459-4313

Chittoor, H. H. S., Griffin, P. R., Neufeld, A., Thompson, J. and Gu, M.
QuLTSF: Long-Term Time Series Forecasting with Quantum Machine Learning.
DOI: 10.5220/0013395500003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 824-829
ISBN: 978-989-758-737-5; ISSN: 2184-433X

ing classical neural networks and variational quantum circuits (VQC's) are increasingly becoming popular for various machine learning tasks, thanks to rapidly evolving software tools (Bergholm et al., 2018; Broughton et al., 2020). The existing QML models for TSF focus on RNN's (Emmanoulopoulos and Dimoska, 2022; Ceschini et al., 2022). In time series analysis, recurrent quantum circuits have demonstrated provable computational and memory advantage during inference, however learning such models remains challenging at scale (Binder et al., 2018).

In this paper we initiate the application of QML to LTSF by proposing *QuLTSF* a simple hybrid QML model. QuLTSF is a combination of classical linear neural networks and VQC's. Through extensive experiments, on the widely used weather dataset, we show that the QuLTSF model outperforms the state-of-the-art linear models proposed in (Zeng et al., 2023).

Organization of the Paper: Section 2 provides background information on quantum computing and quantum machine learning. Section 3 discusses the problem formulation and related work. In Section 4 we introduce our QuLTSF model. Experimental details, results, discussion and future research directions are given in Section 5 and Section 6 concludes the paper.

2 BACKGROUND

2.1 Quantum Computing

The fundamental unit of quantum information and computing is the qubit. In contrast to a classical bit, which exist in either 0 or 1 state, the state of a qubit can be 0 or 1 or a superposition of both. In the Dirac's ket notation the state of a qubit is given as a two-dimensional amplitude vector

$$|\psi\rangle = \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix} = \alpha_0|0\rangle + \alpha_1|1\rangle,$$

where α_0 and α_1 are complex numbers satisfying the unitary norm condition, i.e., $|\alpha_0|^2 + |\alpha_1|^2 = 1$. The state of a qubit is only accessible through measurements. A measurement in the computational basis collapses the state $|\psi\rangle$ to a classical bit $x \in \{0,1\}$ with probability $|\alpha_x|^2$. A qubit can be transformed from one state to another via reversible unitary operations also known as quantum gates (Nielsen and Chuang, 2010).

Shor's algorithm (Shor, 1994) and Grover's algorithm (Grover, 1996) revolutionized quantum algorithms research by providing theoretical quantum speedups compared to classical algorithms. The implementation of these algorithms require larger number of qubits with good error correction (Lidar and Brun, 2013). However, the current available quantum devices are far from this and often referred to as the noisy intermediate-scale quantum (NISQ) devices (Preskill, 2018). Quantum machine learning (QML) is an emerging field to make best use of NISQ devices (Schuld and Petruccione, 2021; Simeone, 2022).

2.2 Quantum Machine Learning

The most common QML paradigm refers to a two step methodology consisting of a *variational quantum circuit* (VQC) or *ansatz* and a classical optimizer, where VQC is composed of parametrized quantum gates and fixed entangling gates. The classical data is first encoded into a quantum state, using a suitable data embedding procedure like amplitude embedding, angle embedding etc. Then, the VQC applies a parametrized unitary operation which can be controlled by altering its parameters. The output of the quantum circuit is given by measuring the qubits. The parameters of the VQC are optimized using classical optimization tools to minimize a predefined loss function. Often VQC's are paired with classical neural networks, creating hybrid QML models. Several packages, for instance (Bergholm et al., 2018; Broughton et al., 2020), provide software tools to efficiently compute gradients for these hybrid QML models.

3 PRELIMINARIES

3.1 Problem Formulation

Consider a multivariate time series dataset with M variates. Let L be the size of the look back window or sequence length and T be the size of the forecast window or prediction length. Given data at L time stamps $\mathbf{x}_{1:L} = \{\mathbf{x}_1, \cdots, \mathbf{x}_L\} \in \mathbb{R}^{L\times M}$, we would like to predict the data at future T time stamps $\hat{\mathbf{x}}_{L+1:L+T} = \{\hat{\mathbf{x}}_{L+1}, \cdots, \hat{\mathbf{x}}_{L+T}\} \in \mathbb{R}^{T\times M}$ using a QML model. Predicting more than 48 future time steps is typically considered as Long-term time series forecasting (LTSF) (Zhou et al., 2021). We consider the channel-independence condition where each of the univariate time series data at L time stamps $\mathbf{x}^m_{1:L} = \{x^m_1, \cdots, x^m_L\} \in \mathbb{R}^{L\times 1}$ is fed separately into the model to predict the data at future T time stamps $\hat{\mathbf{x}}^m_{L+1:L+T} = \{\hat{x}^m_{L+1}, \cdots, \hat{x}^m_{L+T}\} \in \mathbb{R}^{T\times 1}$, where $m \in \{1, \cdots, M\}$. To measure discrepancy between the ground truth and prediction, we use Mean Squared Error (MSE) loss function defined as

$$\text{MSE} = \mathbb{E}_{\mathbf{x}}\left[\frac{1}{M}\sum_{m=1}^{M}\|\mathbf{x}^m_{L+1:L+T} - \hat{\mathbf{x}}^m_{L+1:L+T}\|_2^2\right]. \quad (1)$$

3.2 Related Work

LTSF is an extensive area of research. In this section, we provide a concise overview of works most relevant to our problem formulation. Since the introduction of transformers (Vaswani, 2017), there has been an increase in transformer based models for LTSF (Wu et al., 2021; Zhou et al., 2021; Liu et al., 2022; Zhou et al., 2022; Li et al., 2019). (Wen et al., 2023) provided a comprehensive survey on transformer based LTSF models. Despite the sophisticated architecture of transformer based LTSF models, (Zeng et al., 2023) showed that simple linear models can achieve superior performance compared to the state-of-the-art transformer based LTSF models.

(Zeng et al., 2023) proposed three models: Linear, NLinear and DLinear. Linear model is just a one layer linear neural network. In NLinear, the last value of the input is subtracted before being passed through the linear layer and the subtracted part is added back to the output. DLinear first decomposes the time series into trend, by a moving average kernel, and seasonal components which is a famous method in time series forecasting (Hamilton, 2020) and is extensively used in the literature (Wu et al., 2021; Zhou et al., 2022; Zeng et al., 2023). Two similar but distinct linear models are trained for trend and seasonal components. Adding the outputs of these two models gives the final prediction. We adapt the simple Linear model to propose our QML model in the next section.

4 QuLTSF

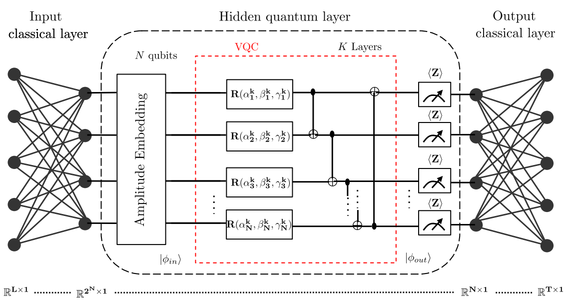

Figure 1: QuLTSF model architecture.

In this section, we propose *QuLTSF* a hybrid QML model for LTSF and is illustrated in Fig. 1. It is a hybrid model consisting of input classical layer, hidden quantum layer and an output classical layer. The input classical layer maps the L input features in to a 2^N length vector. Specifically, the input sequence $\mathbf{x}^m_{1:L} \in \mathbb{R}^{L\times 1}$ is given to the input classical layer with trainable weights $\mathbf{W}_{in} \in \mathbf{R}^{2^N \times L}$ and bias $\mathbf{b}_{in} \in \mathbf{R}^{2^N \times 1}$, and it outputs a 2^N length vector

$$\mathbf{y}_1 = \mathbf{W}_{in}\mathbf{x}^m_{1:L} + b_{in}. \quad (2)$$

The output of the input classical layer, $y_1 \in \mathbf{R}^{2^N \times 1}$, is given as input to the hidden quantum layer which consists of N qubits. We use amplitude embedding (Schuld and Petruccione, 2021) to encode 2^N real numbers in y_1 to a quantum state $|\phi_{in}\rangle$. We use hardware efficient ansatz (Simeone, 2022) as a VQC, and is composed of K layers each containing a trainable parametrized single qubit gate on each qubit and a fixed circular entangling circuit with CNOT gates as shown in Fig. 1. Every single qubit gate has 3 parameters and the total number of parameters in K layers is $3NK$. The output of VQC is given as

$$|\phi_{out}\rangle = (\text{VQC})|\phi_{in}\rangle. \quad (3)$$

We consider the expectation value of Pauli-Z observable for each qubit, which serves as the output of hidden quantum layer and is denoted as $y_2 \in \mathbb{R}^{N\times 1}$. Finally, y_2 is passed through output classical layer with trainable weights $\mathbf{W}_{out} \in \mathbf{R}^{T\times N}$ and bias $\mathbf{b}_{out} \in \mathbf{R}^{T\times 1}$, which maps N length quantum hidden layer output to predicted T length vector

$$\hat{\mathbf{x}}^m_{L+1:L+T} = \mathbf{W}_{out}\mathbf{y}_2 + b_{out}. \quad (4)$$

The parameters of the hidden quantum layer and two classical layers can be jointly trained, similar to classical machine learning, using software packages like PennyLane (Bergholm et al., 2018).

5 EXPERIMENTS

In this section, we validate the superiority of the proposed QuLTSF model through extensive experiments. The code for experiments is publicly available on GitHub[1]. All experiments are conducted on SMU's Crimson GPU cluster[2].

5.1 Dataset Description

We evaluate the performance of our proposed QuLTSF model on the widely used Weather dataset. It

[1] https://github.com/chariharasuthan/QuLTSF
[2] https://violet.scis.dev/

Table 1: Multivariate long-term time series forecasting (LTSF) results in terms of MSE and MAE between the proposed QuLTSF model and the state-of-the-art on the widely used weather dataset. Sequence length $L = 336$ and prediction length $T \in \{96, 192, 336, 720\}$. The best results are in **bold** and the second best results are underlined.

Methods	QuLTSF*		Linear		NLinear		DLinear		FEDformer		Autoformer		Informer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	**0.156**	**0.211**	0.176	0.236	0.182	0.232	0.176	0.237	0.217	0.296	0.266	0.336	0.300	0.384
192	**0.199**	**0.253**	0.218	0.276	0.225	0.269	0.220	0.282	0.276	0.336	0.307	0.367	0.598	0.544
336	**0.248**	**0.296**	0.262	0.312	0.271	0.301	0.265	0.319	0.339	0.380	0.359	0.395	0.578	0.523
720	**0.315**	**0.346**	0.326	0.365	0.338	0.348	0.323	0.362	0.403	0.428	0.419	0.428	1.059	0.741

*QuLTSF is implemented by us; Other results are from (Zeng et al., 2023).

is recorded by Max-Planck Institute of Biogeochemistry[3] and consists of 21 meteorological and environmental features like air temperature, humidity, carbon dioxide concentration in parts per million etc. This is recorded in 2020 with granularity of 10 minutes and contains 52,696 timestamps. 70 percent of the available data is used for training, 20 percent for testing and the remaining data for validation.

5.2 Baselines

We choose all three state-of-the-art linear models namely Linear, NLinear and DLinear proposed in (Zeng et al., 2023) as the main baselines. We also consider a few transformer based LTSF models FEDformer (Zhou et al., 2022), Autoformer (Wu et al., 2021), Informer (Zhou et al., 2021) as other baselines. Moreover (Wu et al., 2021; Zhou et al., 2021) showed that transformer based models outperform traditional statistical models like ARIMA (Box et al., 2015) and other deep learning based models like LSTM (Bai et al., 2018) and DeepAR (Salinas et al., 2020), thus we do not include them in our baselines.

5.3 Evaluation Metrics

Following common practice, in the state-of-the-art we use MSE (1) and Mean Absolute Error (MAE) (5) as evaluation metrics

$$\text{MAE} = \mathbb{E}_{\mathbf{x}}\left[\frac{1}{M}\sum_{m=1}^{M}||\mathbf{x}^{m}_{L+1:L+T} - \hat{\mathbf{x}}^{m}_{L+1:L+T}||_1\right]. \quad (5)$$

5.4 Hyperparameters

The number of qubits $N = 10$ and number of VQC layers $K = 3$ in the hidden quantum layer. Adam optimizer (Kingma and Ba, 2017) is used to train the model in order to minimize the MSE over the training set. Batch size is 16 and initial learning rate is 0.0001. For more hyperparameters refer to our code.

[3]https://www.bgc-jena.mpg.de/wetter/

5.5 Results

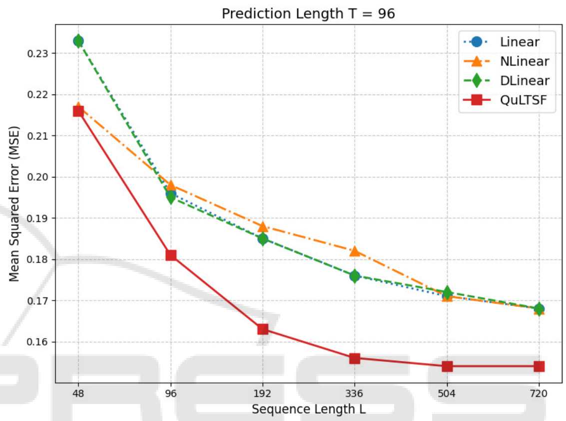

Figure 2: MSE comparison with fixed prediction length $T = 96$ and varying sequence length $L \in \{48, 96, 192, 336, 504, 720\}$.

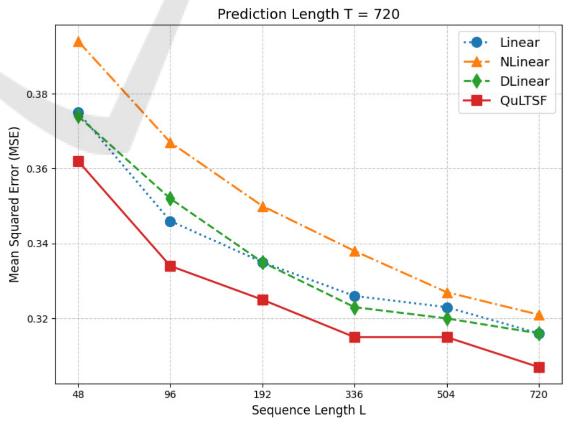

Figure 3: MSE comparison with fixed prediction length $T = 720$ and varying sequence length $L \in \{48, 96, 192, 336, 504, 720\}$.

For a fair comparison, we choose the fixed sequence length $L = 336$ and 4 different prediction lengths $T \in \{96, 192, 336, 720\}$ as in (Zeng et al., 2023; Zhou et al., 2022; Wu et al., 2021; Zhou et al., 2021). Table 1 provides comparison of MSE and MAE of QuLTSF with 6 baselines. The best and second best results are highlighted in **bold** and underlined respectively. Our proposed QuLTSF outperform all the baseline models in all 4 cases.

To further validate QuLTSF against the baseline linear models we conduct experiments for the QuLTSF and classical models with varying sequence lengths $L \in \{48, 96, 192, 336, 504, 720\}$ and plot the MSE results for a fixed smaller prediction length $T = 96$ in Fig. 2, and for a fixed larger prediction length $T = 720$ in Fig. 3. In all cases QuLTSF outperforms all the baseline linear models.

5.6 Discussion and Future Work

QuLTSF uses generic hardware-efficient ansatz. Similar to classical machine learning in QML we need to choose the ansatz, if possible, based on dataset and domain expertise. Searching for optimal ansatz is a research direction by itself (Du et al., 2022). Finding better ansätze for QML based LTSF models for different datasets is an open problem. One potential way is to use parameterized two qubit rotation gates (You et al., 2021).

Other possible future direction is to use efficient data preprocessing, for example reverse instance normalization (Kim et al., 2021) to mitigate the distribution shift between training and testing data. This is already being used in the state-of-the-art transformer based LTSF models like PatchTST (Nie et al., 2023) and MTST (Zhang et al., 2024). These models also show better performance than the linear models in (Zeng et al., 2023). Interestingly, our simple QuLTSF model outperforms or comparable to these models in limited settings. For instance, for the setting ($L = 336, T = 720$) MSE of PatchTST, MTST and QuLTSF are 0.320, 0.319 and **0.315** respectively. For the setting ($L = 336, T = 336$) MSE of PatchTST, MTST and QuLTSF are 0.249, **0.246** and 0.248 respectively (see Table 2 in (Zhang et al., 2024) for PatchTST and MTST; and Table 1 for QuLTSF). QML based LTSF models with efficient data preprocessing may lead to improved results.

Implementation of QML based LTSF models on the quantum hardware poses significant challenges. As quantum systems grow, maintaining qubit state coherence and minimizing noise become increasingly difficult, leading to errors that degrade computational accuracy and performance (Lidar and Brun, 2013). Additionally, the barren plateau phenomenon, where the gradient of the cost function vanishes as system size or circuit depth increases, further complicates optimization of VQC's (McClean et al., 2018). Addressing these challenges requires innovations in error correction, problem-specific circuit designs, and alternative optimization strategies, all of which are critical for enabling scalable, noise-resilient, and effective QML based LTSF models.

6 CONCLUSIONS

We proposed *QuLTSF*, a simple hybrid QML model for LTSF problems. QuLTSF combines the power of VQC's with classical linear neural networks to form an efficient LTSF model. Although simple linear models outperform more complex transformer-based LTSF approaches, incorporating a hidden quantum layer yielded additional improvements. This is demonstrated by extensive experiments on a widely used weather dataset showing QuLTSF's superiority over the state-of-the-art classical linear models. This opens up a new direction of applying hybrid QML models for future LTSF research.

ACKNOWLEDGMENTS

This research is sponsored by the Singapore Quantum Engineering Programme (QEP), project ID : NRF2021-QEP2-02-P06.

REFERENCES

Bai, S., Kolter, J. Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M. S., Alonso-Linaje, G., Akash-Narayanan, B., Asadi, A., et al. (2018). Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*.

Binder, F. C., Thompson, J., and Gu, M. (2018). Practical unitary simulator for non-markovian complex processes. *Physical review letters*, 120(24):240502.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Broughton, M., Verdon, G., McCourt, T., Martinez, A. J., Yoo, J. H., Isakov, S. V., Massey, P., Halavati, R., Niu, M. Y., Zlokapa, A., et al. (2020). Tensorflow quantum: A software framework for quantum machine learning. *arXiv preprint arXiv:2003.02989*.

Ceschini, A., Rosato, A., and Panella, M. (2022). Hybrid quantum-classical recurrent neural networks for time series prediction. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Du, Y., Huang, T., You, S., Hsieh, M.-H., and Tao, D. (2022). Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*, 8(1):62.

Emmanoulopoulos, D. and Dimoska, S. (2022). Quantum machine learning in finance: Time series forecasting. *arXiv preprint arXiv:2202.00599*.

Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hyndman, R. (2018). *Forecasting: principles and practice*. OTexts.

Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., and Choo, J. (2021). Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.

Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32.

Lidar, D. A. and Brun, T. A. (2013). *Quantum error correction*. Cambridge university press.

Lim, B. and Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209.

Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A. X., and Dustdar, S. (2022). Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*.

McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., and Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*.

Nielsen, M. A. and Chuang, I. L. (2010). *Quantum computation and quantum information*. Cambridge university press.

Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum*, 2:79.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3):1181–1191.

Schuld, M. and Petruccione, F. (2021). *Machine learning with quantum computers*, volume 676. Springer.

Shor, P. (1994). Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134.

Simeone, O. (2022). An introduction to quantum machine learning for engineers. *Found. Trends Signal Process.*, 16(1–2):1–223.

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems. Neural information processing systems foundation*, page 5999.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L. (2023). Transformers in time series: a survey. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6778–6786.

Wu, H., Xu, J., Wang, J., and Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430.

You, J.-B., Koh, D. E., Kong, J. F., Ding, W.-J., Png, C. E., and Wu, L. (2021). Exploring variational quantum eigensolver ansatzes for the long-range xy model. *arXiv preprint arXiv:2109.00288*.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. (2023). Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128.

Zhang, Y., Ma, L., Pal, S., Zhang, Y., and Coates, M. (2024). Multi-resolution time-series transformer for long-term forecasting. In *International Conference on Artificial Intelligence and Statistics*, pages 4222–4230. PMLR.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115.

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. (2022). Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pages 27268–27286. PMLR.