№082

ICQT 2025

VIII INTERNATIONAL CONFERENCE ON QUANTUM TECHNOLOGIES

Network Traffic Analysis with Quantum Machine Learning

S. Manko[1], A. Moiseevskiy[1,2]

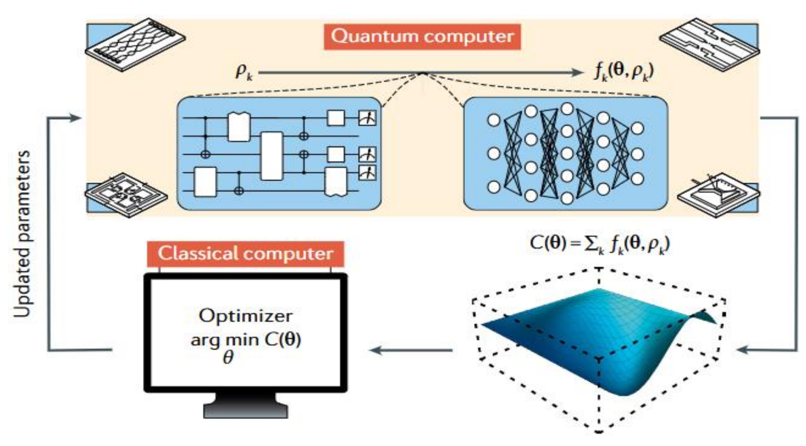
[1]InfoTeCS Scientific Research and Advanced Developments Centre

[2]MSU Quantum Technology Centre

e-mail: *Sofia.Manko@infotecs.ru*

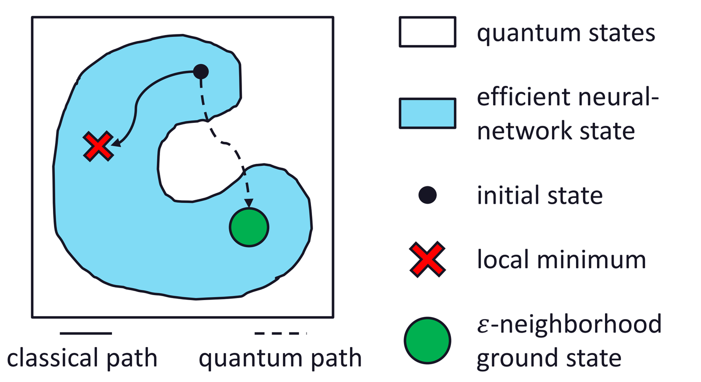
Network traffic prediction using historical data is crucial for maintaining performance, preventing congestion, and detecting anomalies or cyber threats. This work explores the application of quantum machine learning (QML) for network traffic forecasting. It is assumed that QML models have a natural ability to identify periodic dependencies in data, since parameterized quantum circuits can easily be represented as Fourier series. However, current QML results for time-series prediction do not provide a clear picture of the effectiveness of applying quantum models to this type of problem.

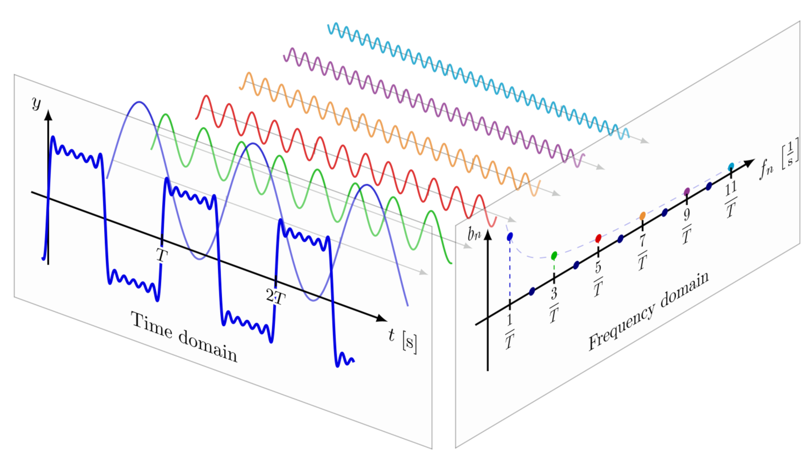
Introduction



Variational quantum algorithms combine parameterized a quantum circuit with a classical optimizer. Being more noise-resistant [1], they could find practical applications earlier than other quantum algorithms requiring extremely low error rates, driving active research into their NISQ-era applications.

It is believed that due to the atypical patterns generated by quantum systems in Hilbert space, **quantum advantage** may be achieved through shorter paths to the optimum [2].





Parameterized quantum circuits can be naturally represented using Fourier series:

$$x = (x_1, \dots, x_N) \qquad \theta = (\theta_1, \dots, \theta_N)$$

$$f_\theta(x) = \sum_{n \in \Omega} c_n(\omega) e^{inx}$$

Which means that they could be promising choice for **time series forecasting** [3].

Methods

SARIMA is a statistical analysis model consisting of the following components:

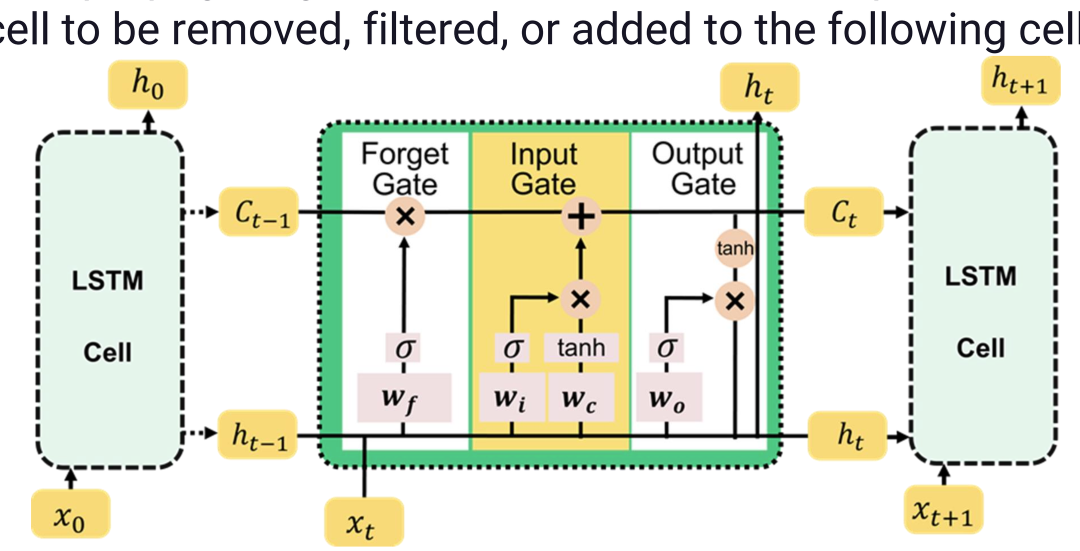
(S) Seasonal: The seasonal period is user-defined based on known data characteristics

(AR) Autoregressive: Measures the correlation between the current value and the past ones: $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$

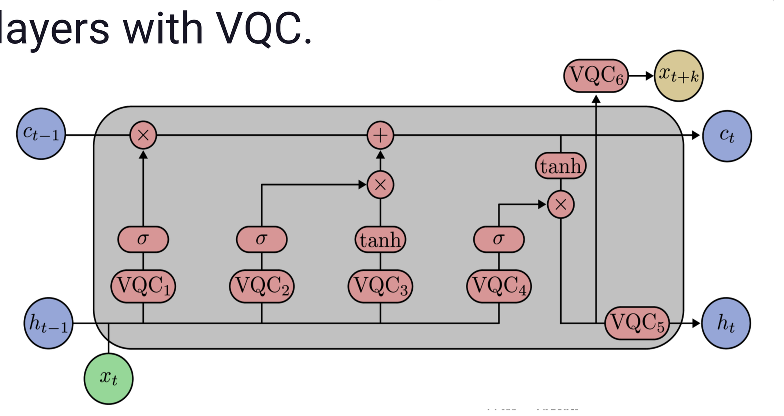
(I) Integrated: Represents the number of times the series needs to be differenced to make it stationary, so that it can be described by an ARMA model

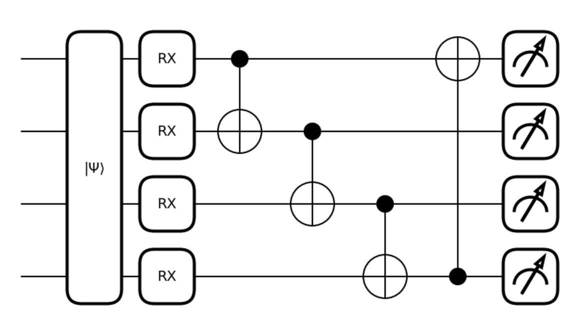
(MA) Moving Average: Measures the correlation between the current value and the past error terms of the AR model: $y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots$

LSTM (Long Short-Term Memory) is a specialized recurrent neural network that uses memory cells to capture and store sequential data. Cells link past and present modules, propagating information across time steps. Internal gates allow data in each cell to be removed, filtered, or added to the following cells. [PyTorch]

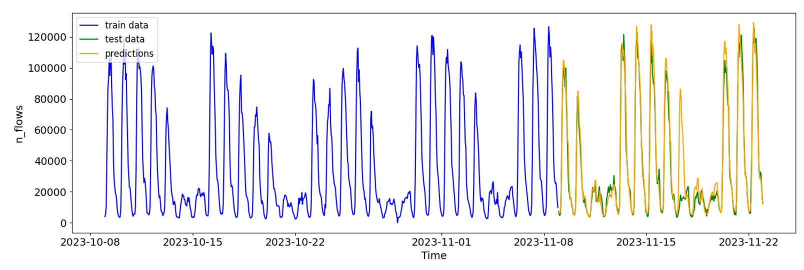


QLSTM is a hybrid quantum-classical model that combines the advantages of LSTM and variational quantum circuits (VQC) is constructed by replacing classical layers with VQC.

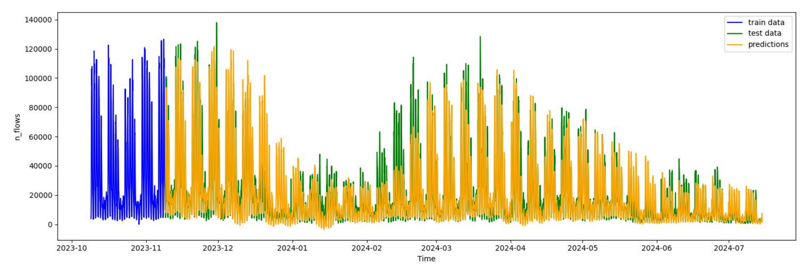




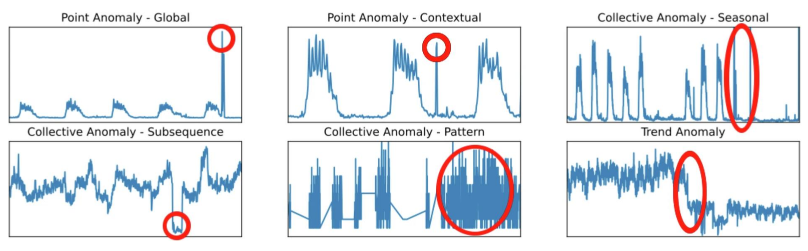
Data



Example of model forecasting (train 744 hours (31 d) / test 335 h (14 d))



Example of forecasting with retraining (train 744 h (31 d) / test 168 h (7d) / retrain interval (7 d))



Examples of typical anomalies present in the dataset [5]

Rising network traffic demands accurate **forecasting** to optimize resources and detect threats. While ML outperforms statistical methods for complex patterns [4], it faces scalability challenges with big data.

In this work we used a real network traffic **dataset** containing information about the Czech Interuniversity Network [5]. For prediction based on historical data, one feature was chosen - the number of flows.

Another important task is **anomaly detection** in time series, which can be performed through methods such as threshold exceedance of the difference between predicted and actual values or using autoencoders [6], though this remains a subject for future research.

Results

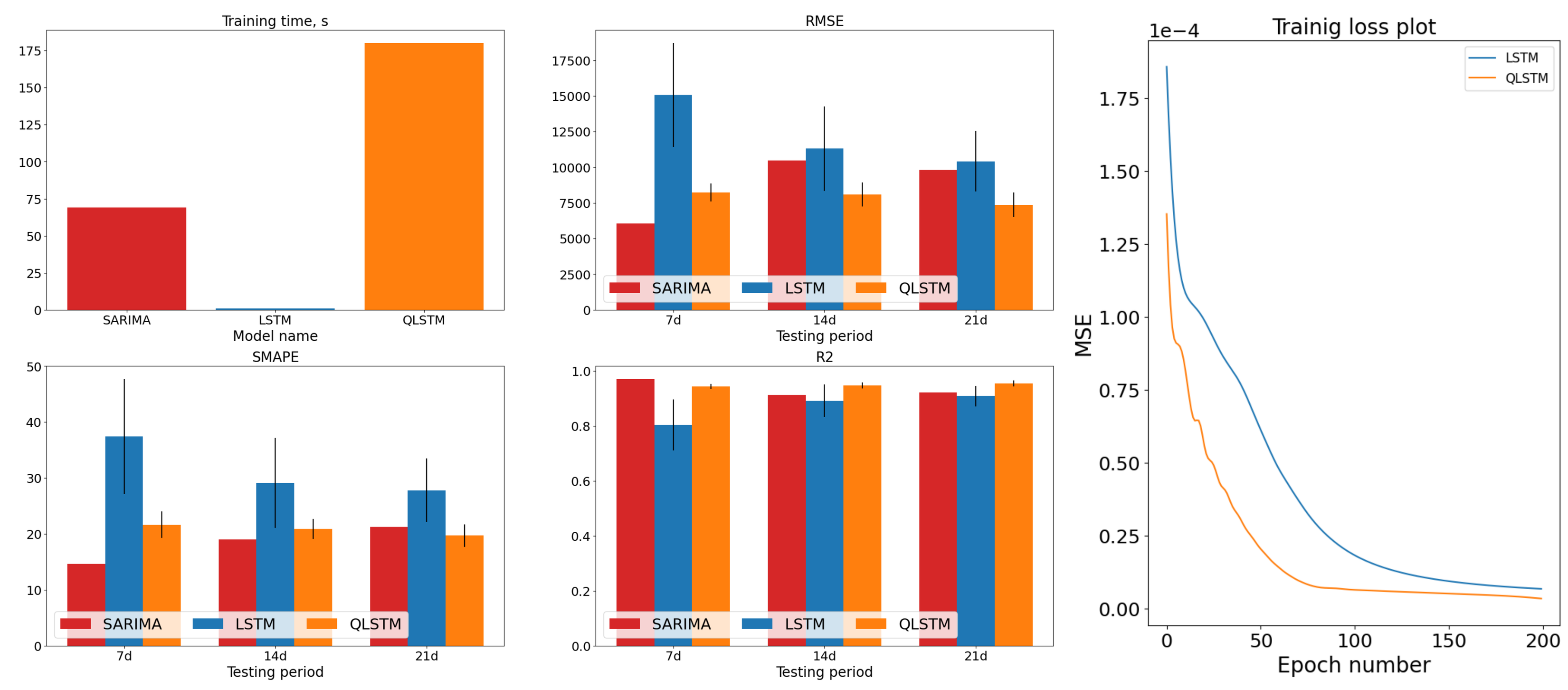
The metrics used to evaluate the performance of the models are the following:

$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ - differences between predicted and actual values

$SMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}$ - symmetric percentage error metric

$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ - proportion of explained variance

Evaluated model metrics:



Based on the presented results, SARIMA outperforms on highly stationary data segments, while QLSTM shows more stable run-to-run training and higher accuracy than LSTM.

Further research plans: realization of accurate rolling forecasting via all models, use of hardware-efficient ansatz to enable the experiment

Visit GitHub for technical details and future research updates:

References

[1] K. Sharma *et al.*, New J. Phys. **22**, 043006 (2020).

[2] J. Biamonte *et al.*, Nature **549**, 195 (2017).

[3] M. Schuld, R. Sweke, J. J. Meyer, Phys. Rev. A **103**, 032430 (2021).

[4] O. Aouedi *et al.*, ACM Comput. Surv. **57**, 1(2025).

[5] J. Koumar *et al.*, Sci. Data **12**, 338 (2025).

[6] R. Frehner, K. Stockinger, Quant. Mach. Intell. **7**, 1 (2025).

[7] https://github.com/Reflex-Angle/QLSTM-Prediction/tree/main