NATURE MACHINE INTELLIGENCE INITIAL SUBMISSION FORMAT

Hasnain Sikora, 19BAI1072

ABSTRACT

As the decades have passed so has the boundaries of technological advancement; the world we live in has witnessed the accretion of ubiquitous data that directly or indirectly has helped make better business decisions and research preparedness around multiple sectors. With abundance of data, it has become imperative for individuals to make accurate predictions of the future based on historical datasets. This is where time-series data lands. Time-series forecasting has since become a rather common task wherein historical data and current data points are utilized to make informed predictions over a period of time.

Pattern recognition, Resource allocation, Sports analytics Statistics, and Weather forecasting are only a few of the most critical applications of taking advantage of past predictions stored in the memory to predict future behaviour accurately. Time-series data can be extrapolated using machine learning methods like random forests, gradient boosting, and time delay neural networks using rolling mean statistics. The inherent functionality of a good time-series undoubtedly gouges information by recognizing sequences outside a supplied training data in contrast to how traditional machine learning methodologies work. This is where the notion of LSTMS and Quantum LSTMS come in. Quantum technology itself provides a boom in the future of computation.

Quantum computers act as a medium of super-function by allowing information bits to exist as a superposition of multiple states at the same time. With the advent of multiple algorithms that allow these computers to process a glut of data parallelly computing results , data entanglement and their ability to share parameters and process multiple qubits together has allowed these systems to solve computation-intensive problems such as decryption and quantum simulations exponentially faster than a traditional system. The combination of these systems with classical data being converted into a wave function and applying the principles of coherence and decoherence along with quantum operations foments a gateway for us to replicate the function of an artificial neural network which would in theory perform efficiently on problems where the dataset is surfeit. The algorithms which work to allow the function of a Quantum Neural Network hybrid have shown promising results in performance when put through data classification and computer vision problems. To distend upon the efficacy of QNNs, we use Quantum-LSTMS applied to time-series datasets, contending as promising tools in various industries such as finance, healthcare, and manufacturing.

INTRODUCTION

LSTMs were first introduced in 1997 by Hochreiter and Schmidhuber, and have since become a popular tool for a extensive gamut of time series forecasting functions [1]. [2]. The key advantage of LSTMs over traditional RNNs is their ability to selectively remember past information for long periods of time [3]. This is achieved through the use of gating mechanisms, which allow the network to essentially add constraints to the information that flows through the model's hidden state. One of the most popular applications of LSTM is time-series forecasting, whose functionality lies fundamentally in prognosticating possible values of a time series data-sequence based on its past values. Time series forecasting is broadly used in a glut of fields, including financial sequences [4], economics [5], and weather forecasting [6]. LSTMs have been shown to be highly effective for this task, outperforming traditional methods such as ARIMA and exponential smoothing [7]. Since we our article is focused on QLSTMS, we must first introduce QML whose primary aim is essentially to leverage properties of quantum mechanics, such as superposition as well as entanglement, to ameliorate the performance of deep learning algorithms [8]. This permits quantum computers to parallelly execute calculations, which potentially is a testament to their speed as compared to classical computers [9].

Recently, there has been significant accretion in the development of quantum computing not only with regards to hardware but also the construction and execution of algorithms on these platforms [10]. Quantum computing, as mentioned, performs executions and calculations with regards to qubits, which can essentially manipulate information in a non-binary state due to the property of superposition [11]. Quantum operations can be executed on qubits when exposed to quantum circuits, which are essentially a sequential set of operations that change the state of these qubits. They do so by transmuting their relative phases in the Hilbert space. The state of qubits is widely delineated by a Bloch sphere, and operations on qubits correspond to rotations of the quantum state vector in this implied space [12]. The fundamental goal of Quantum Machine Learning (QML) is to substitute certain components of a neural network, such as linear layers, with their quantum counterparts [13].

Quantum LSTMs (QLSTMs) are a recent development that aim to leverage the properties of quantum mechanics to improve the performance of traditional LSTMs. QLSTMs also have the ability to perform certain types of computations that are not possible with traditional LSTMs, such as quantum machine learning and quantum optimization. It's worth noting that the research in this field is ongoing and still in its early stages, and it's not yet clear whether QLSTMs will be able to deliver on their potential benefits. However, research work with regards to Quantum neural networks describe a promising future.

In [14], author explores the idea of using the Hilbert space for NLP by taking advantage of the fact that NLP is rooted in classical statistics. Quantum mechanics, is essentially defined by the artithmetic of quantum statistics, which distend upon traditional statistics by delineating objects by matrices of complex numbers. The author implements a quantum LSTM using a Variational Quantum Circuit (VQC) which is a type of quantum circuit that utilizes rotation angles to optimize a cost function. Nonetheless, a VQC cannot individually alter the spatial dimensionality of the input, so the author loads it between two classical linear layers to match the desired dimensionality. The author achieved a comparable if not better accuracy when compared to the classical LSTM. The variational circuit block consisted of tunable parameters and its output was then extracted back into its classical form to be fed into the corresponding forget gates. The model achieves 99% accuracy with a 0.056 loss which is comparable to a classic LSTM but uses less than half of the parameters being fed into that model.

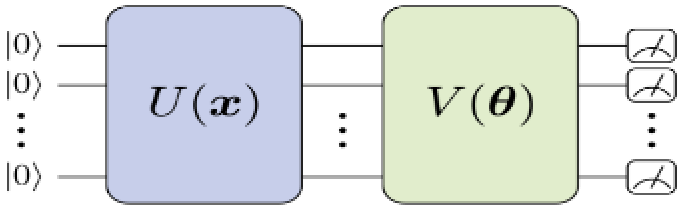


Figure 1. Architecture Of A Generic VQC replacing the nodes in a classical LSTM

In [15], the author proposes a model that performs better than a traditional GRU and LSTM by introducing a transformer encoder with an attention head. The dataset used cleverly uses temporal factors such as hour-number, and daythe of year along with their cosine and sine distributions to factor in contrasting values of TEC between the first and last hour of the day. The model uses 120 hours of data from 5 days to predict the 6th day. The proposed system also compares itself to the GRU, which combines the input and forget gates into an update gate, and has only an update and reset gate. The update gate uses historical information, while the reset gate regulates meaningful information transmission. However, due to the limitations of GRU, a transformer mechanism was introduced. The author’s transformer mechanism is essentially an encoder-decoder architecture which contains multiple transformer blocks with feed-forward and multi-head attention. The purpose of multi head attention is to enhance diversity by focusing on information from various representation subspaces at varying positions. This gives us a look at how parallel computation of data can be worked out in non-quantumized scenarios.

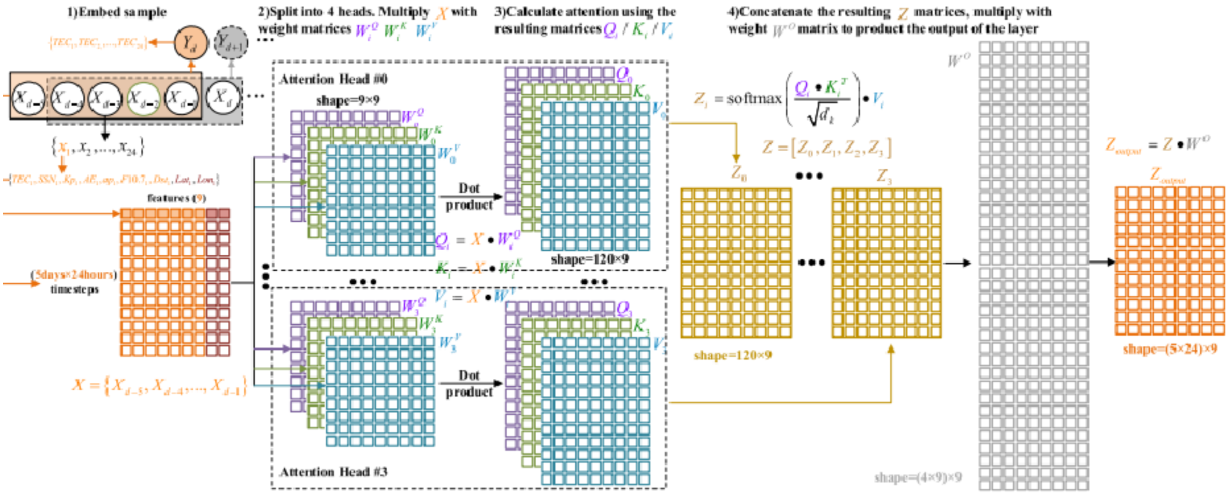

1)Embed sample
2)Split into 4 heads. Multiply X with weight matrices W_i^Q W_i^K W_i^V
3)Calculate attention using the resulting matrices Q_i / K_i / V_i
4)Concatenate the resulting Z matrices, multiply with weight W^O matrix to product the output of the layer
features (9)
(5days×24hours) timesteps
Attention Head #0
shape=9×9
Dot product
shape=120×9
Attention Head #3
Z_i = softmax(Q_i • K_i^T / √d_k) • V_i
Z = [Z_0, Z_1, Z_2, Z_3]
shape=120×9
W^O
shape=(4×9)×9
Z_output = Z • W^O
shape=(5×24)×9
X = {X_{d-5}, X_{d-4}, ..., X_{d-1}}

Figure 2. Architecture of an attention based transformer model which takes in a feature vector for each of the 5 days

In [16], the author introduces a detailed working of QNN and the multiple quantum algorithms used in order to make effective QNNs. The DQNN takes care of the issue of state-collapse by introducing a unitary operation to solve the issue of the mixed states collapsing into one another but also applies Deutsch's algorithm which essentially introduces the qubit systems as a tensor-product that disables the unitary operation from collapsing the pure states. The DQNN takes into consideration the previous layers of quantum perceptron much like a classical LSTM and has the added benefit of lesser memory requirements and the ability of train deep networks.

In [17], the authors introduce an Bidirectional QLSTM for language identification in code-mixing between hindi and english. The proposed architecture follows creating an encoded system of the words using word2vec libraries which are done in a modular approach. The first module consists of the MNN processes and hot-encodes the words to derive word and character embeddings which are useful when extracting the meaning, POS and N-gram of the words present. This information is then passed to the Q-LSTM encoding layer. This layer is where the roman-script based character embedding of hindi texts and its variations are used in the memory to generate text-labels for the input string. The model achieved a relatively high f-score when computing the similarity score of the hindi-roman words and through it, a similarity score was given to the ones that consisted of a variation of that word.

In [18], the author introduces recurrent quantum neural networks. The proposed Quantum Recurrent Neural Network successfully performs consequential tasks like digit classification and data-sequence training. The QRNN cell is made of parametrized quantum neurons and uses amplitude amplification to activate polynomial inputs and cell states, leading to a nonlinear activation and the ability to predict class probabilities at each step. The RQNN consists of a variational quantum circuit made using ancillas that have rotational gates and quantum-controlled gates to orchestrate the state of the qubits.

The output for the cell is a distributed probability calculated to transform the quantum result to a classical format to be fed to the next cell. The authors applied this to a sequence memorization and generative task by using different input vectors consisting of one or two digits along with the full fledged MNIST dataset. With data augmentation applied the QRNN performed consistently well by giving an accuracy of 95%.

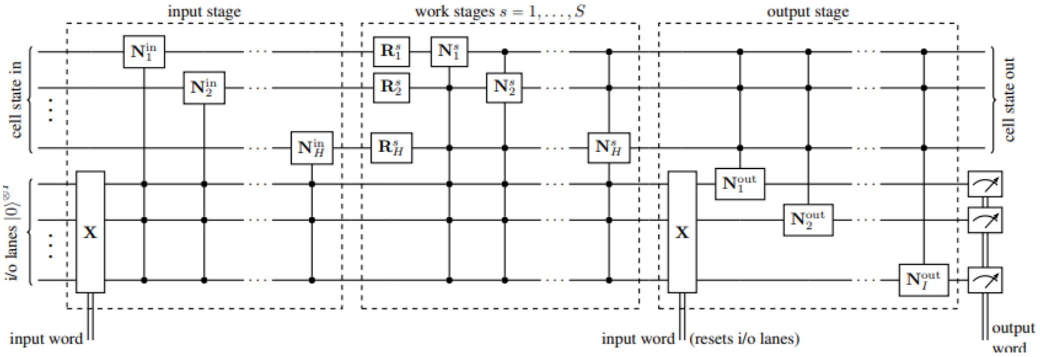


Figure 4. Quantum recurrent neural network cell

In [19], the authors introduce a hybrid QLSTM architecture which effectively learns various types of time-based data and converges faster and achieves accuracy a cut above than its traditional counterpart. The VQCs mentioned are essentially sets of quantum gates that have adaptive parameters which is exactly what the system needs when trying to update the parameters during error correction much like classical neural networks that back-propagate to tune its weights. The U(X) block is an encoding block that prepares classical data for quantum operations. Hadamard gate is used which are operated on phase-based vectors which can be multiplied with the density matrix of the qubit to achieve superposition. The V(x) block is analogous to traditional networks as they take in theta which are parameters that can be optimized using gradient based learning. Since the QLSTM is essentially a hybrid, the model also uses a measurement block which is a probability distribution function that computes the probability with which the state will collapse. This gives a resulting classical output that can be then fed into the corresponding gate for further processing.

In [16]. the author introduces a highly portable quantum language model (PQLM) that can transmit information to classical machines for downstream problems with ease This model, built with random Variational Quantum Classifiers (VQC) and localized models, demonstrates commensurate performant abilities to its traditional counterpart both intrinsically and extrinsically with regards to its evaluation metrics. The authors also study the factors affecting PQLM performance and provide a theoretical foundation for future NLP research with privacy protection demands. As quantum computing hardware becomes more widely available, the authors envision their approach could be used to train large-scale PQLMs and easily transmit the quantum information to classical tasks.

In a recent study [20], the author proposed a hybrid time series forecasting model called the Fuzzy-Quantum Time Series Forecasting Model (FQTSFM) that addresses the two primary issues namely the selection of the universe of discourse and the determination of the fuzzy degree of memberships. To tackle these problems, the author integrated the Fast Forward Quantum Optimization Algorithm (FFQOA) with the FTS modeling approach. The FQTSFM model was tested with three time series datasets consisting of various properties, including daily average temperatures of Taipei, Taiwan Futures Exchange (TAIFEX) index, and Taiwan Stock Exchange Corporation (TSEC) weighted index. Results from various statistical metrics confirmed the superior forecast accuracy of the FQTSFM compared to other established FTS and non-FTS models. The model proposed was concluded to be much efficient with regards to its CPU's time utilization. The constraint of this research was that it was corroborated only with univariate time series set, however, the model can be distended to handle multivariate time series.

The author introduces a novel LSTM based recurrent architecture for time series forecasting in [21]. The architecture tensorizes the cell's state-to-state propagation in LSTM, which enhances the learning of short term nonlinear complexity while retaining the LSTM's memory feature. The efficiency and generality of the architecture were investigated experimentally throughout the article. The author takes into account two many-body entanglement structures, matrix product states (MPS) and a multiscale entanglement renormalization ansatz (MERA), as tensor decomposition techniques, and found that the proposed MERA network performs better than the MPS, suggesting that the learnability of the system's entropy is concluded by not only the free parameters but also the complexity of the tensor. The experiments revealed that LSTM-MERA yielded better results than the LSTM-MPS without the requirement of altering the number of parameters, leading to the conjecture that the tensor's entanglement entropy has to grow with the system's size for efficient learning of chaos.

In a groundbreaking study by Emmanoulopoulos and Dimoska [22], the use of parametrized quantum circuits (PQCs) as quantum neural networks (QNNs) was explored for time series forecasting. The performant abilities of PQCs was put into comparison against a classical BiLSTM using temporal signals consisting of a combination of sinusoidal components, trends, and additive noise. The results showed that for signals with low amplitude noise variations (up to 40% of the amplitude of the deterministic signal), PQCs with few parameters performed similarly to BiLSTM networks with thousands of parameters and outperformed them for signals with higher noise variations.

TITLE	MODELS	DATASET	RESULTS	PROS	CONS
LSTM-Designed Quantum Experiments	LSTM	MELVIN generations	Model performs well even on validation data proving its inference capacity. Generative models can be of service for complex quantum experiments	Generative models can be of service for complex quantum experiments	Nil
FQTSFM: A fuzzy-quantum time series forecasting mode	FQTSFM	TAIFEX index, Taipei's average temperatures and TSEC weighted index	Fast convergence, ability to develop one-step ahead forecast results, improved forecast accuracy compared to existing models, effective with regards to CPU's time utilization	Fast convergence, ability to develop one-step ahead forecast results, improved forecast accuracy compared to existing models, effective with regards to CPU's time utilization	Not tested with multivariate time series datasets
Language identification framework in code-mixed social media text based on quantum LSTM — the word belongs to which language?	BiLSTM	ICON 2016 POS Tagging and FIRE2014 Transliterated Search	A comprehensive analysis of identifying language in a code mixed date, successful evaluation models, and effectiveness of word embeddings and BLSTM system	A comprehensive analysis of identifying language in a code mixed date, successful evaluation models, and effectiveness of word embeddings and BLSTM system	Lack of system scaling to identify code-mixing and language characteristics in data with a mix
QUANTUM LONG SHORT-TERM MEMORY	QLSTM	Different kinds of temporal data	Comparison to classical LSTM shows QLSTM Has a faster convergence rate and attains a higher level of accuracy.	Comparison to classical LSTM shows QLSTM Has a faster convergence rate and attains a higher level of accuracy.	Further research could explore the limitations and limitations of the QLSTM model in different types of temporal data and investigate the potential limitations in implementation on NISQ devices
THE DAWN OF QUANTUM NATURAL LANGUAGE PROCESSING	QLSTM	IMDB dataset for binary classification	QLSTM took 10 min to train, compared to classical LSTM's 10 sec QLSTM needed longer training, 300 epochs, to reach 100% accuracy with a loss of 0.056	First to train a QLSTM for parts-of-speech tagging and the first to propose a quantum-enhanced Transformer for sentiment analysis	Complexity of simulation of quantum circuit took longer than classical counterpart

Table 1: Comparison of related works

RESULTS

To fully realize the potential of the Quantum LSTMS, we drew up comparisons of three different qubit systems to witness a difference in computation power. To simplify the task and keep it constrained to performance capacity, A set of 4-qubits, 8-qubits and 6-qubits systems were trained for 15 epochs each.

In this section, we describe the results of our experiments where we trained and evaluated two different models, LSTM and QLSTM, on our weather stats dataset. We compared the performance of these models using the train and test loss metrics. The LSTM was trained with 20 epochs and another where we trained the QLSTM with 15 epochs on the same series. The LSTM model achieved a train loss of 0.00013 and a test loss of 0.000191. This indicates that a well-defined LSTM was in-line with the task such that it performed forecasts with a minimal error variation. The LSTM and QLSTM models perform similarly, but it is important to notice that the QLSTM converges to the minima faster when than the LSTM. After careful trial and error with various combinations of hyperparameters, the one that were narrowed down for comparison between the two models are listed in Table 2.

Hyperparameter	**QLSTM**	**LSTM**
Learning Rate	0.05	0.0001
No. of Qubits	6	Nil
No. of Epochs	20	20
No. of Hidden Units	16	16
Loss Function	MSE	MSE
Optimizer	Adagrad	Adam

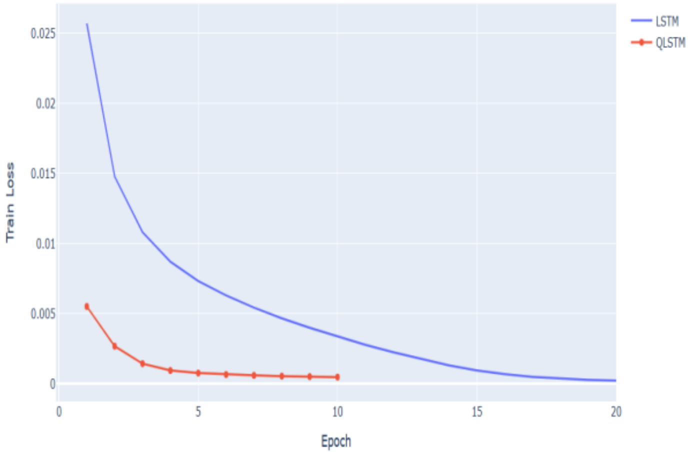
Network parameters for best performing Models

They are able to deftly predict the training set of about 6000 data points and provide highly accurate forecasts for a significant period after the test set begins, before gradually but slightly diverging from the actual values. The QLSTM thus not only converged much faster than the LSTM but did so in half the amount of epochs indicating the leverage of parallelism using qubit operations. As shown by the test reports in [], the QLSTM model achieved a train loss of 0.000468 and a test loss of 0.0004537. As summarized in Table 3 below.

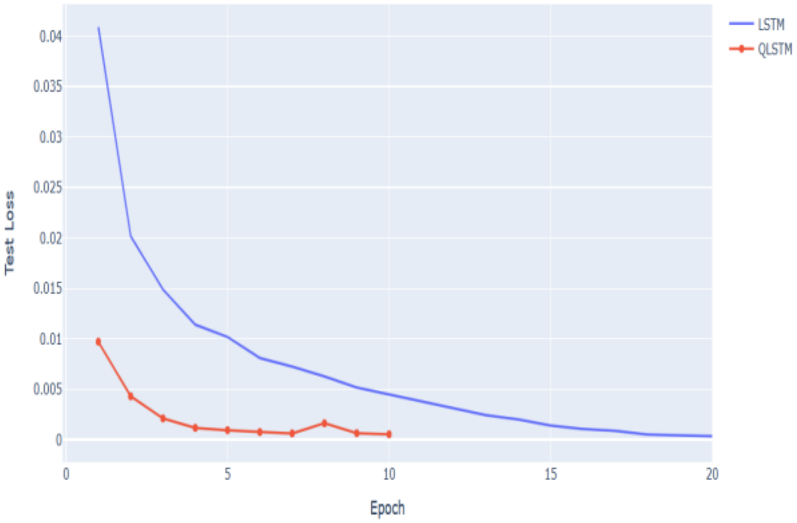
Category	Model	Epochs	Avg. Execution Time per Epoch (s)	Loss (MSE)	
				Train	Test
Classical	LSTM	20	350	0.00013	0.00019
Quantum	QLSTM – 4 Qubits	15	3800	0.000217	0.00028
		20	4000	0.000194	0.000195
	QLSTM – 6 Qubits	15	4200	0.00017	0.000273
		20	5000	0.000144	0.000176
	QLSTM – 8 Qubits	15	6400	**0.000149**	**0.000139**
		20	7000	**0.00075**	**0. 00010**

Comparision of formulated LSTM Models

This indicates the potential of the QLSTM model to outperform LSTM in terms of speed and efficiency. Overall, our study suggests that QLSTM models have the potential to provide better performance and faster convergence than LSTM models, especially when trained on large datasets. We present the results of the experiments where we trained the models with 10 epochs. The train and test loss graphs for the LSTM and QLSTM models are shown in the following figures:

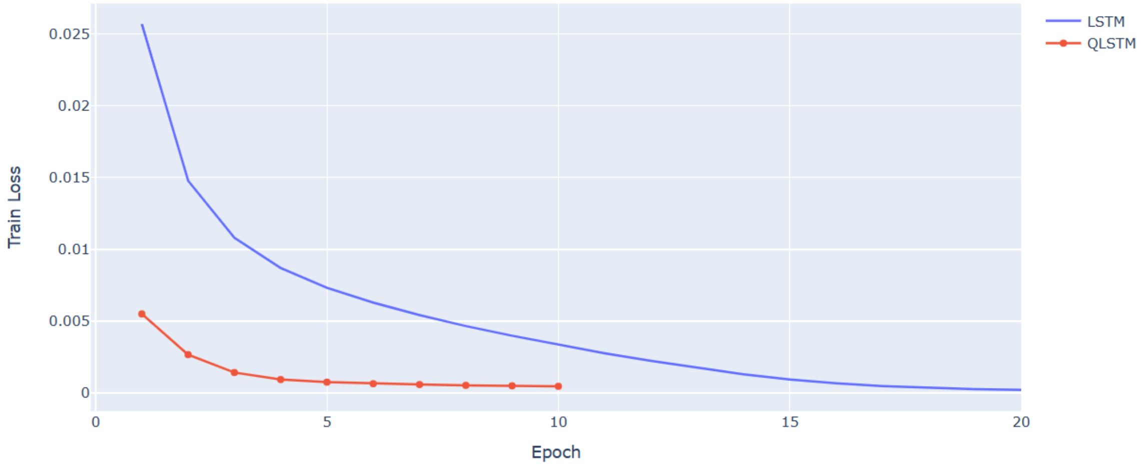


QLSTM VS LSTM Train Loss

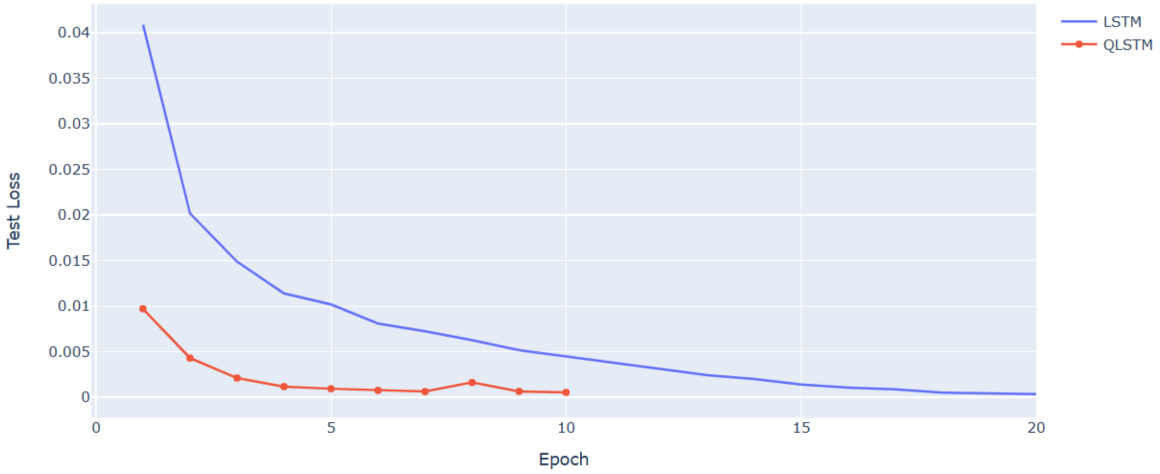


QLSTM VS LSTM Test Loss

We thus present the results for the experiments where we plotted the training loss of the QLSTM and LSTM models on the same graph and the test loss of the QLSTM and LSTM models on the same graph, to signify the faster convergence of the QLSTM as opposed to the classical LSTM and thus uses half the amount of training steps as well as tunable parameters that are used. As expected, multiple qubits, did in fact increase the memory space exponentially and thus used much more time to train however, they did achieve improved accuracy in much lesser epochs all thanks to parallelism [16]. The 4-qubit system, 6-qubit system and an 8-qubits system essentially raises the Hilbert space by C^16.



4-QUBIT QLSTM VS LSTM Train Loss Over 10 epochs

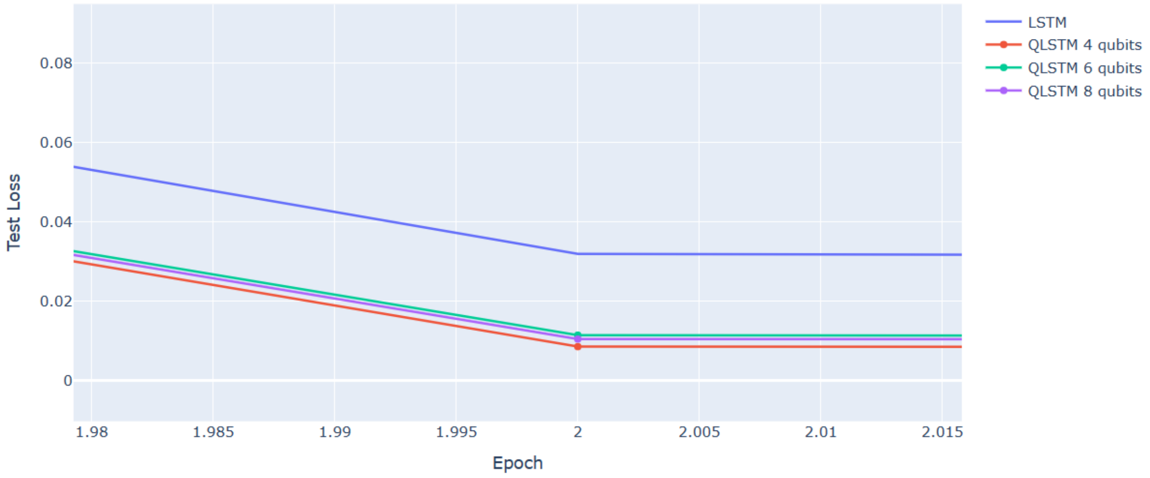


4-QUBIT QLSTM VS LSTM Test Loss Over 10 epochs

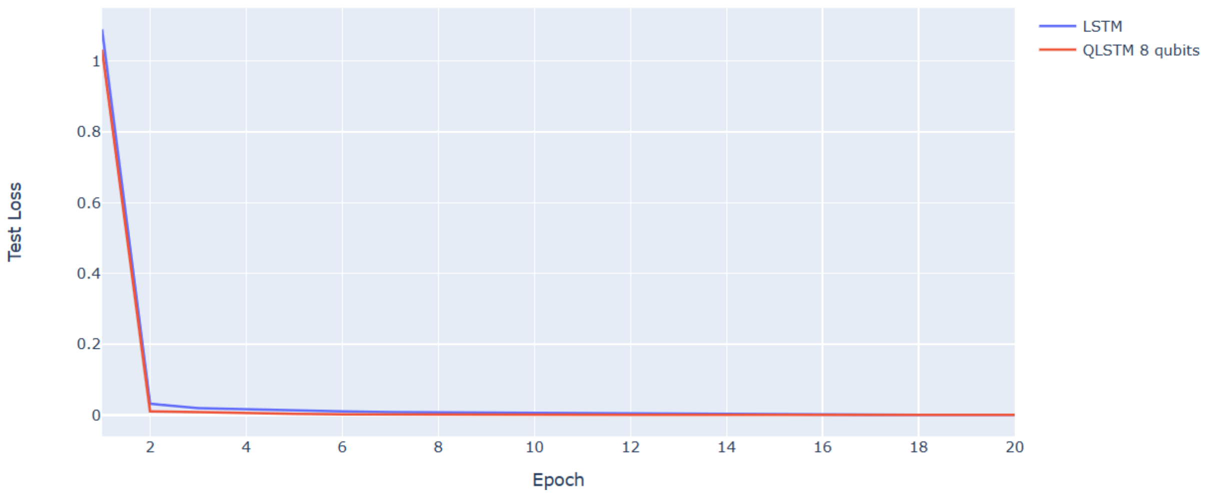
The system was intially trained using a 4-Qubit system trained for half the number of training epochs and as noticed, the QLSTM converged at an exponentially faster rate in its initial steps and by the end of the 10th epoch it subsists with a test loss and train loss of 0.000453 and 0.00046887 respectively. It did witness a turbulent 8th epoch but soon converged further down the training steps. However, for us to understand the true potential of qubit systems, it was time for us to increase the number of epochs as well the number of qubits that the system took into consideration. As mentioned already, this task was going to be extremely computationally taxing because this increase would exponentially increase the Hilbert space by a factor of $2n$.

The 4-Qubit system performed excellently already with over 20 epochs trained, it narrowed down to an test loss of 0.0002832. This 6-Qubit system was also trained with 20 epochs, it narrowed down to a test loss of 0.0002732. The 8-Qubit system undoubtedly converged much faster and bought the test loss down further than our proposed loss to a whopping 0.00010 and performed better than the LSTM.

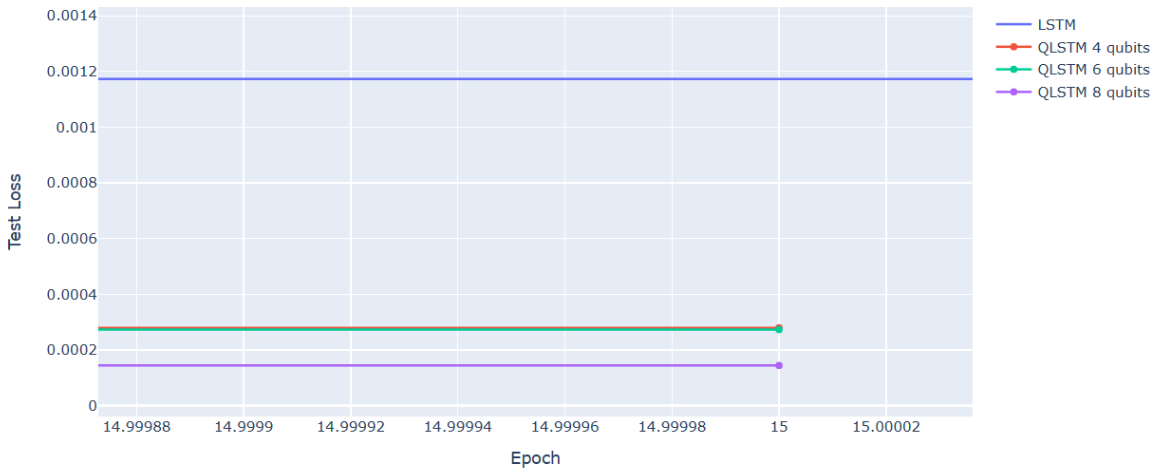
Due to the ability to share its parameters, it has also not only used lesser number of tunable parameters that the VQC intakes but also does so in lesser number of epochs. The following graphs show the QLSTMS and the LSTMs test loss over their respective train steps.



All QLSTMs VS LSTM Test Loss Over 10 epochs Zoomed In



8-Qubit QLSTMs VS LSTM Test Loss Over 20 epochs



All QLSTMs VS LSTM Test Loss Over 15 epochs Zoomed In

To further buttress our findings, we have also taken into consideration an SFA LSTM as referenced in [37.]. Which was trained on a variation of the same Saskatoon weather stats data. The SFA LSTM took into account a learning rate of 0.0001 and 32 hidden units as opposed to 16 units in our specific problem statement. The results are summarized in Table 4 and it can be noted that the QLSTM outperforms the SFA LSTM by not only converging much faster in the given epochs but also resulting in a minimal loss. Similarly, other classical variants such as Stacked, Conv, Seq2seq and Enc-Dec LSTM also follow the same trend and are delineated in the following table.

Model	Learning Rate	Hidden States	Optimizers	Epochs	MSE Test Loss	Reference
BiLSTM	0.01	16	Adam	20	0.136	[37]
SFA LSTM	0.0001	32	Adam	25	0.0871	[38]
Stacked LSTM	0.01	16	Adam	20	0.1106	[41]
ConvLSTM	0.01	16	Adam	20	0.51771	[42]
Seq2seq LSTM	0.01	16	Adam	20	0.2015	[43]
Enc-Dec LSTM	0.01	16	Adam	20	0.139	[44]
QLSTM	**0.05**	**16**	**Adagrad**	**20**	**0.0001**	

A Comparative Analysis of QLSTM and Classical LSTM Variants

There is no suitable contrast if we are unable to include quantum variants that are tasked with a time-series or sequential data prediction task. Since the datasets, are vastly different, we use this opportunity to shed light upon the number of parameters, loss values for their specific datasets, amount of hidden units and so forth. This following section only serves to put our QLSTM tasked with forecasting weather against existing Quantum network variants and not to draw analysis.

Feature	**QLSTM**	**Parametrized Quantum LSTM**
Architecture	Quantum LSTM	Parametrized Quantum LSTM
Simulator	PennyLane’s standard qubit-based device	Qasm Simulator
Shots	1000	100
Learning Rate	0.05	0.001
Hidden Units	16	32
Optimizers	Adagrad	Adam
Epochs	20	5000
Loss	**0.00010**	0.7

QLSTM and PQ-LSTM

The table depicted above essentially sheds light on a modified classical network [36.] . The PQ-LSTM was tasked with a time-series task of monitoring stress on various time-series data such as computer interaction, posture and facial expressions which changed overtime with regards to induced stress actuators.

This specific time-series task, while not suited to be compared with weather forecast gives us a potential application of using quantum variants and how they fare against classical counterparts. The PQ-LSTM converged to a 0.7 negative log loss after a set parameter of 100 quantum circuit shots. Additionally, it was tested upon a much smaller dataset consisting of 3000 datapoints trained to 5000 epochs. This was essential to log as not only does it use the VQC as a pre-processing layer but also applied a different quantum simulator owing to more variation. The table below also sheds light on the other quantum variants focussed on various other problem statements and signifies the performant nature of the QLSTM when it comes to our prediction task. Performing just as good if not better than the other quantum variants.

Model	Problem Statement	Dataset	Loss Metrics	Reference
QMN	The paper proposes a new framework, called QMN, to accurately and comprehensively model complicated interactions in multi-modal sentiment analysis in conversations.	MELD, IEMOCAP	0.72 (F1-Score)	[39]
Q-BiLSTMA	The paper proposes a Q-BiLSTM with attention model for ADR detection, using variable component subcircuits (VQC)	TwiMed	>10 (Cross Entropy)	[40]
PQ-LSTM	A model is proposed to support knowledge workers in managing their stress levels using sensor data obtained from various modalities	HRT	0.87 (Negative Log)	[36]
QLSTM	QLSTM that works on a specific multivariate weather prediction use-case	Weatherstats	0.0001 (MSE)	

DISCUSSION

With almost every new technological advancement in the world, classical computing also has its own caveats [9]. Classical supercomputers are without a doubt unable to describe quantum systems because of the general ability of quantum systems to exponentially increase the Hilbert space that they lie in as the amount of qubits shoot up. This limitation fomented the birth of quantum computers championed by Richard Feynman [10]. These new quantum computers would allow users to simulate a many-body environment as opposed to struggling classical computers. There has also been a glut of key quantum algorithms, including the Deutsch-Jozsa algorithm [12], Shor's factoring algorithm [13], and Grover's search algorithm [9].

With faster systems and computing power as well as the advent of publicly available quantum computers, the processing of quantum information has become experimentally possible [10-13]. These quantum computers, referred to as noisy intermediate-scale quantum (NISQ) devices, typically have a couple hundred qubits and enable researchers to test out quantum algorithms under induced noise levels. While the public of NISQ devices [22] has led to a cornucopia of ventures to solve problems on these devices, it is important to acknowledge that computations on these quantum systems with roughly 100 qubits can more often than not be simulated using classical computers.

When we pull our focus back to QLSTMs which was simulated on a classical computer with a set of 4,6 and 8 qubits, It becomes evident that QLSTM learns significantly more information during the initial epochs compared to LSTM, leading to much faster convergence to an acceptable loss value. This improved trainability of the quantum model is a noteworthy advantage over its classical counterpart. Eventually, both models converge to a low loss value, but the QLSTM does so in a shorter amount of time due to its ability to quickly assimilate information. This project has demonstrated that QLSTM can be a promising approach for time-series prediction, specifically weather forecasting, yielding results comparable to classical LSTM models while requiring fewer parameters for training and achieving greater information gain per epoch.

However, there is still room for further improvement in this technique, specifically the need for a real quantum system with a powerful processing power capable of handling more epochs that would converge to a better-trained model with proprietary quantum computation instead of the Pennylane simulation that is being used currently to showcase the potential of a QLSTM For Time Series Prediction. The future scope of QLSTMs is vast and holds great promise. They have the potential to revolutionize fields such as cryptography, speech and text generation as well as of course forecasting tasks.

However, these promising systems are not without their limitations. In addition to those mentioned above, one of the biggest challenges in the development of Quantum-LSTMs is their sensitivity to noise and other sources of error, which can lead to a loss of coherence also known as state collapse and thus result in loss of information. This makes it difficult to maintain the stability of qubits, which are the basis of quantum computers. With continued research and development, Quantum neural networks may one day engender a new generation of computing and data processing that will transform the way we approach artificial intelligence.

METHODS

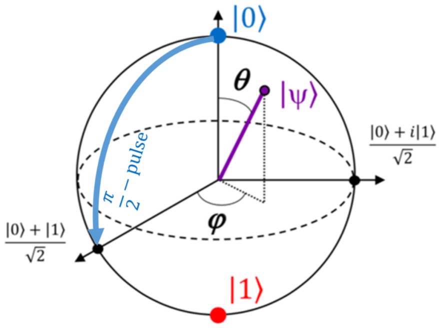
Quantum computers work a little differently than classical computers. Albeit they are also dependent on fundamental information bits, they are called qubits and they differ in the sense that they represent a two-state quantum system and are denoted by something known as the Ket notation [8, 25]. Ket denotations allow us to represent our systems as extremities in various systems which makes it easier for us to process tasks related to complex physical systems. Qubits have the ability to be coherent. they are in a state of superposition wherein the qubit can subsist as both 0 and 1 which are denoted using the Dirac notation as - $|0\rangle$ and $|1\rangle$. Mathematically we can describe the qubits by the following equation:

$$|0\rangle = (1\ 0)\quad |1\rangle = (0\ 1)\ (i)$$

A qubit that subsists in a so-called pure-state can be delineated by |φ> which is a normalized vector in a hypothetical two-dimensional Hilbert space. This can be portrayed by the following mathematical equation:

$$|\psi\rangle \ = \ \alpha\,|0\rangle \ + \ \beta\,|1\rangle \ \textit{(ii)}$$

In this equation, α and β are complex numbers whose sum of squares equates to 1. They are both probability amplitudes whose squares represent the probability of the qubit leaning toward the corresponding state.



$|0\rangle$
θ
$|\psi\rangle$
$\frac{|0\rangle + i|1\rangle}{\sqrt{2}}$
$\frac{\pi}{2} - pulse$
φ
$\frac{|0\rangle + |1\rangle}{\sqrt{2}}$
$|1\rangle$

$$|\psi\rangle = cos\frac{\theta}{2}|0\rangle + e^{i\varphi}sin\frac{\theta}{2}|1\rangle$$

A Bloch sphere

The equation of a pure state can be written as follows to represent them on a Bloch sphere:

$$|\psi\rangle \ = \ cos(\theta/2)|0\rangle \ + \ e\text{^}(i\phi)sin\left(\frac{\theta}{2}\right)|1\rangle \ \textit{(iii)}$$

where θ and ϕ are phases that determine the location of a point on the Bloch sphere. The angle θ represents the inclination of the point from the positive z-axis and is given by: $\theta \ = \ 2cos^{-1}(\alpha)$ The angle ϕ represents the azimuthal angle of the point around the z-axis and is given by: $\phi \ = \ 2\pi arg(\beta)$ where arg(β) is the argument of β, which equates to the complex number’s phase.

$$p0 \ = \ \langle\psi|0\rangle\,\langle 0|\psi\rangle \ = \ |\langle 0|(\alpha\,|0\rangle \ + \ \beta\,|1\rangle)^2 \ = \ |\alpha < 0|0 > |^2 = \ |\alpha|^2 \ \textit{(iv)}$$

How do we procure a value to process in our applications since we cannot just process qubits? Similar to classical bits, a qubit can have two possible outcomes when measured, often denoted as 0 or 1. When a qubit is measured in the basis described by equation above, it collapses into one of the pure states, either |0⟩ or |1⟩. It becomes possible to calculate the probabilities of the collapsing by obtaining a specific measurement outcome. For example, the probability of measuring 0 is given by p0 = $|\alpha|^2$, where α is a coefficient in the qubit state vector. The measurement outcomes are associated with measurement operators, which are represented by |0⟩⟨0| ,|1⟩⟨1| respectively [26]. The complex numbers are thus probability amplitudes that measure the probability towards which the state collapses.

MULTI-QUBIT SYSTEMS

Akin to classical bits, which demonstrate their full capability when employed in long chains of bits, the potential of a quantum system, including quantum computing, is best realized when multiple qubits are employed. We delineate an n-qubit with a tensor product of 2^{n} basic states subsisting in the Hilbert space denoted as $H = \left(C^{2}\right)^{\otimes n} = C^{2n}$. Mathematically this can be represented as the following:

$$|q1\rangle \otimes |q2\rangle \otimes ... \otimes |qn\rangle = |q1q2 ... qn\rangle \ (v)$$

A simple pragmatic example would be to take a two-qubit system. We will be using a 4-qubit, 6-qubit and 8-qubit to draw parallels in our QLSTMS. For the time being however, let us focus on how we can represent a two-qubit system. The fundamental states of a 2-qubit system can be represented as $|00\rangle, |01\rangle, |10\rangle\ and\ |11\rangle$. A pure two-qubit state can be described as a linear combination or superposition of these states, as shown in the equation:

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle \ (vi)$$

QUANTUM ENTANGLEMENT

Entanglement is a fundamental phenomenon in quantum algorithms., It requires at least two qubits [16, 20] to demonstrate. A quantum system is said to be entangled, if it cannot be separated. A pure n subsystem is deemed separable if it can be depicted in the following form

$$|\psi\rangle = |\psi 1\rangle \otimes \cdots \otimes |\psi n\rangle \ (vii)$$

Consider a two-qubit system of qubits, A and B. The state of qubit A is given by : $|\psi A\rangle = \alpha|0\rangle + \beta|1\rangle$ And the state of qubit B is given by: $|\psi B\rangle = \gamma|0\rangle + \delta|1\rangle$. We can write the composite system of those two qubits as:$|\psi AB\rangle = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle$. This would be a separable state as they can be represented as follows:

$$|\psi AB\rangle = (\alpha|0\rangle \otimes \gamma|0\rangle) + (\alpha|0\rangle \otimes \delta|1\rangle) + (\beta|1\rangle \otimes \gamma|0\rangle) + (\beta|1\rangle \otimes \delta|1\rangle) \ (viii)$$

However, when we take another two qubits system represented by the following equation: $|\psi\rangle = \frac{(|00\rangle + |11\rangle)}{\sqrt{2}}$ To prove that they are not entanglement they must satiate the following equation.

$$|\psi\rangle = |\alpha\rangle \otimes |\beta\rangle \quad (ix)$$

However, as we distend upon the above equation, we find that upon expanding a generic two qubit system as follows:

$$|\psi\rangle = ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle \quad (x)$$

It will never equate to the state equation that we have mentioned above as no values of variables a, b, c or d subsist that would satiate the equation. This means that the two-qubit system we proposed is entangled.

QUANTUM GATES

Quantum circuits are essential in utilizing quantum mechanics for quantum computing as well as creating a full fledge quantum block for our QLSTM. They are composed of quantum operations and are divided into two subordinate categories – one-qubit gates and multi-qubit gates.

Quantum gates are operations used in quantum computing, similar to logic gates in classical computing. Some classical gates have direct quantum analogues, such as the NOT gate. The NOT gate, which transfers the state |0⟩ to |1⟩ and |1⟩ to |0⟩ in classical computing, can also be represented in Dirac notation as the Pauli X-gate. It takes the form:

$$X = |0\rangle\langle 1| + |1\rangle\langle 0| \quad (xi)$$

Applying this gate to our pure states of either $|0\rangle \, or \, |1\rangle$ does the quantum operation as described in the following equations:

$$X\,|1\rangle = |1\rangle\,\langle 0|1\rangle + |0\rangle\,\langle 1|1\rangle = |0\rangle \quad (xii)$$

$$X\,|0\rangle = |1\rangle\,\langle 0|0\rangle + |0\rangle\,\langle 1|0\rangle = |1\rangle \quad (xiii)$$

The X-gate is only one of the many one-qubit gate that performs the NOT operation, which swaps the states |0⟩ and |1⟩. The quantum gates are constrained by the requirement that they preserve the normalization of quantum states, which means that the sum of the squared magnitudes of the coefficients of the state vector must equal 1.

Unitary gates are a class of quantum gates that satisfy this constraint, and they are frequently used in many quantum algorithms. Some other commonly used one-qubit gates include the Hadamard gate, which can create superpositions of states and is often used for measuring qubits in specific bases.

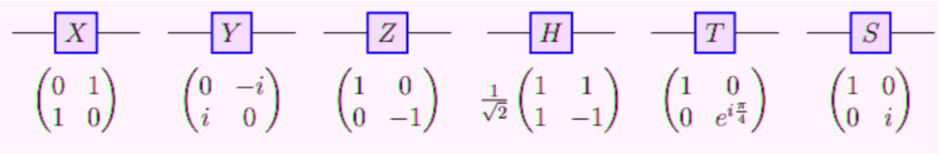

X
Y
Z
H
T
S
$\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$
$\begin{pmatrix}0 & -i\\i & 0\end{pmatrix}$
$\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$
$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}$
$\begin{pmatrix}1 & 0\\0 & e^{i\frac{\pi}{4}}\end{pmatrix}$
$\begin{pmatrix}1 & 0\\0 & i\end{pmatrix}$

Figure 16. Common Quantum Gates

Another common gate is the CNOT gate is a two-qubit gate in quantum computation that performs a NOT operation on the target qubit (the second qubit) if the control qubit (the first qubit) is in the state $|1\rangle$. In any other case, the target qubit stays unchanged. Let's consider an example where the control qubit is in the state $|0\rangle$ and the target qubit is in the state $|1\rangle$. The CNOT gate acts on this two-qubit state as described the following equation:

$$cnot\,|01\rangle = |01\rangle \quad (xxiii)$$

In this case, the control qubit is in the state $|0\rangle$, so the target qubit stays unchanged. Now let's consider an example where the control qubit is in the state $|1\rangle$ and the target qubit is in the state $|0\rangle$. The CNOT gate acts on this two-qubit state as follows:

$$cnot\,|10\rangle = |11\rangle\ (\ SEQ\ (\backslash\ *\ \ roman\ xxiv)$$

In this case, the control qubit is in the state $|1\rangle$, so the target qubit sustains a NOT operation and is flipped from $|0\rangle\ or\ |1\rangle$. These examples demonstrate how the CNOT gate allows for conditional operations to be performed on two qubits in a quantum circuit.

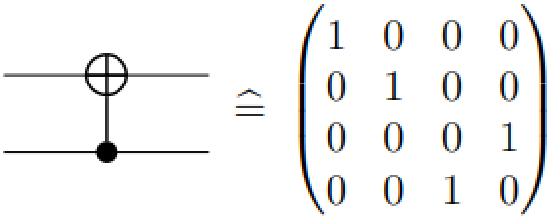


Figure 17. CNOT Gate Denotion And Matrix Form

QUANTUM CIRCUITS

There is only a limited quantity of gates that can be directly implemented in a quantum computer, so we need to find ways to approximate complex operations using sequences of elementary quantum gates much akin to digital logic gates. The Solovay-Kitaev theorem mentions that any unitary gate can be realized with a certain level of accuracy using a sequence of gates from a finite set. Some sets of gates, called universal sets, can be used to build any single-qubit gate.

One commonly used set contains the gates CNOT, H, S, and T [27]. Quantum circuits are a way to represent quantum operations using gates and other operations like tensor products and measurements.

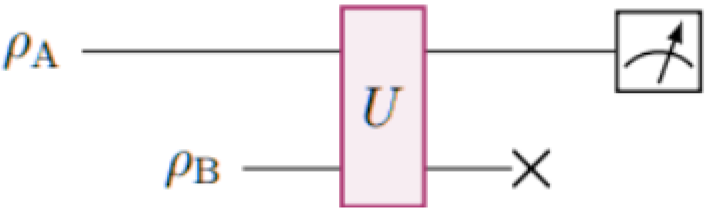


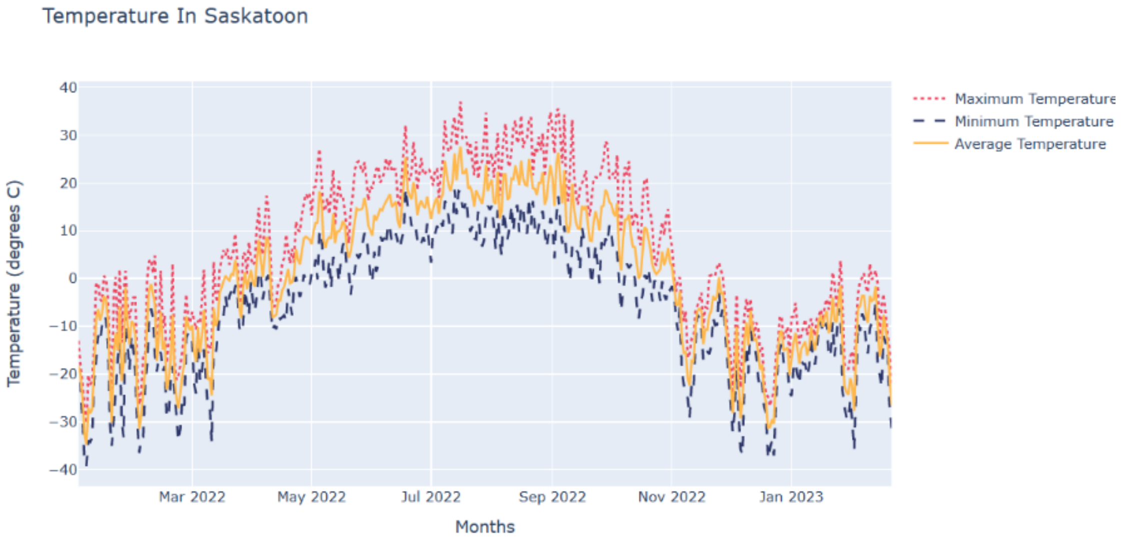
Figure 18. Basic Unitary Gate Implemented Into A Quantum Circuit

The above diagram represents a basic unitary gate implemented into a quantum circuit which takes in two density matrices and thus gives us an classical measurement by calculating the probability with which it collapses into a resulting state. This acts as a basis of unitary state operations that allow us to build upon to construct our QLSTM

This section dives a little deeper into the application and technical side of our problem statement. The dataset was also pre-processed and trimmed down to only take into consideration the most significant features. The following section distends upon the dataset selection.

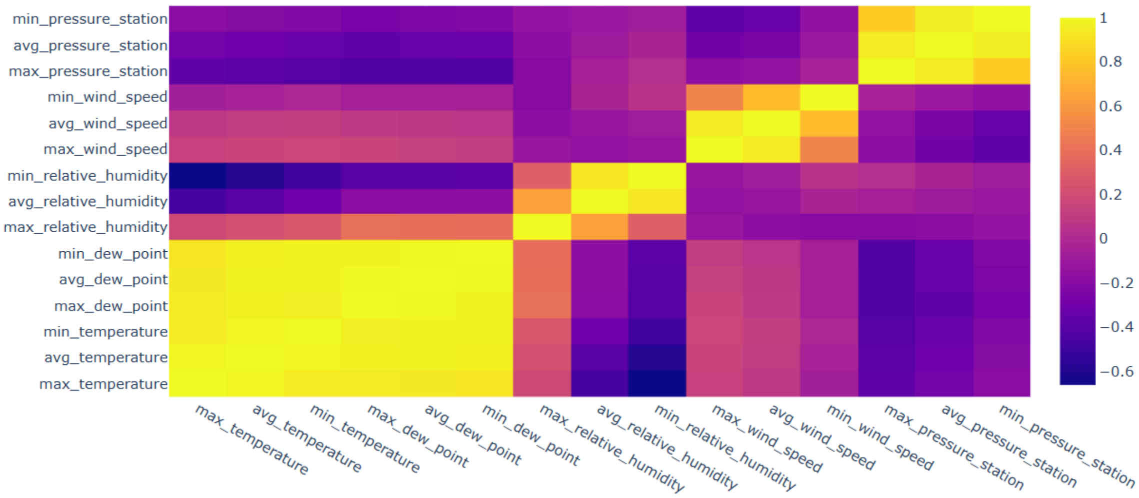
DATASET DESCRIPTION

The 30-year normals (daily) on https://saskatoon.weatherstats.ca refer to a set of climatological data collected over a 30-year period. These data represent the typical weather conditions for a given location, and they are used to provide a baseline for comparison with current or future weather conditions. The data are collected from the Environment and Climate Change Canada climate station located in Saskatoon, Saskatchewan, Canada.



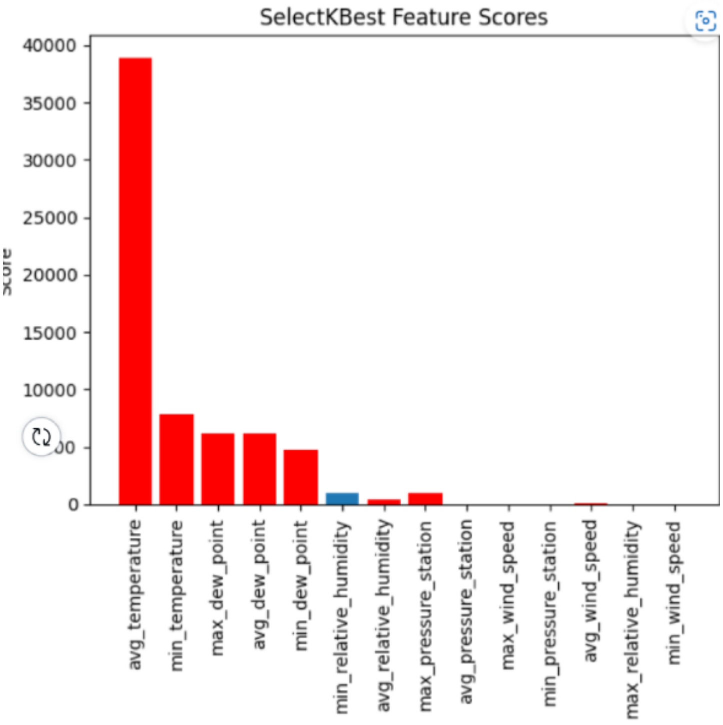
Graph of target variable

This set of data would provide a good beginning point to explore and divide our data in a way that could do justice to the training capabilities of both LSTM and our QLSTM. The crux of the problem statement is to accomplish time-series weather forecasting using a quantum LSTM. The dataset is collected from the weather statistics available on the Weather Stats website, providing actual meteorological data from the weather station located at Saskatoon John G. Diefenbaker Intl. Airport, with a latitude of 52.14 and longitude of -106.69. The Climate Daily/Forecast/Sun Saskatoon weather dataset from weather stats contains historical weather data for the city of Saskatoon, Saskatchewan, Canada.



Correlation heatmap of dataset features

It includes daily observations of various meteorological parameters, such as temperature, relative humidity, dew point, station pressure, windchill, wind speed and sea pressure. These data points as mentioned before may not all be important as the amount of correlation is varied. This is where a famous bio-evolutionary method comes in, genetic algorithms, as well as SelectKBest methodology, are used from the scikit-learn library to select the most important features from our dataset.



Top selected features

The method specifies the value of k as 12, indicating that the top 12 features will be selected. It is initialized with the f-regression score function which computes the correlation between each feature and the target variable and returns the corresponding F-value and p-value for each feature. The method then selects the top k features based on the highest F-values. [6].

The code defines a PyTorch module for a quantum long short-term memory (QLSTM) network. The QLSTM module processes sequences of input data, where each element of the sequence is a feature vector of fixed length. The QLSTM network uses quantum circuits to perform computations at each time step. It has several hyperparameters, including the size of the input and hidden layers, the number of qubits in the quantum circuits, the number of quantum layers, and the number of variational rotations in each quantum layer.

The module also supports different backend devices for quantum computations, which can be specified at initialization. The module defines four quantum circuits that correspond to the four gates in a standard LSTM network: forget, input, update, and output. Each circuit takes a batch of input features and a set of trainable weights as inputs and outputs a vector of expectation values for Pauli-Z operators applied to each qubit in the circuit [32]. The quantum circuits are implemented using the PennyLane library, and the weights are learned using backpropagation through the quantum circuits. LSTM is a layer of chained gates that are responsible for memorizing and forgetting information coupled with the linear networked layer.

This is exactly where a hybrid layer can be sandwiched such that the VQC is divided into layers that encode, train and give out a measured probability distribution that is then passed to a corresponding sigmoid or tanh function to be fit in the memory and passed to the next layer. Libraries such as PennyLane or TensorFlow Quantum offer a high level of abstraction to define quantum/digital hybrid models for Quantum Machine Learning.

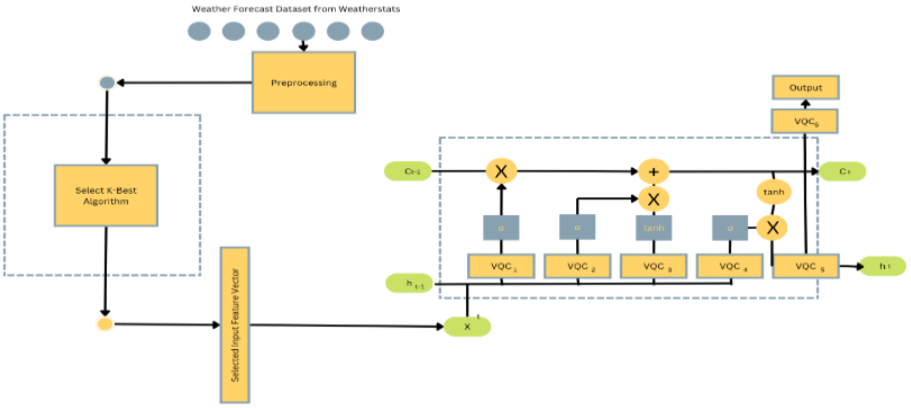


Figure 19. QLSTM System Architecture

The architecture of our Quantum LSTM follows a variational quantum circuit as mentioned above. The base methodology is the same as it remains in the working of the LSTM. The only major difference is what happens inside the variational quantum circuit. A VQC essentially is a block of unitary operations which pre-processes our classical data and essentially prepares the state [13, 33].

Then it passes through the variational quantum circuit which essentially is analogous to a traditional neural network in the sense that it allows the updating of parameters and improve the accuracy. As mentioned in the sections above, it becomes easy for us to understand the working of the working of these gates and the circuit as the whole.

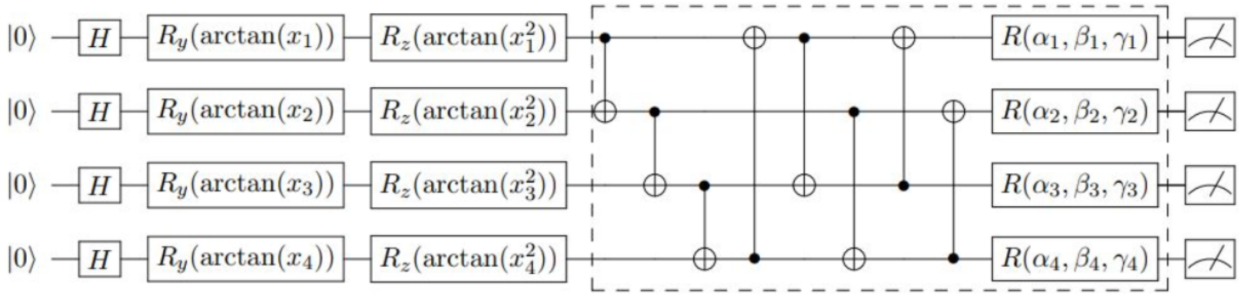
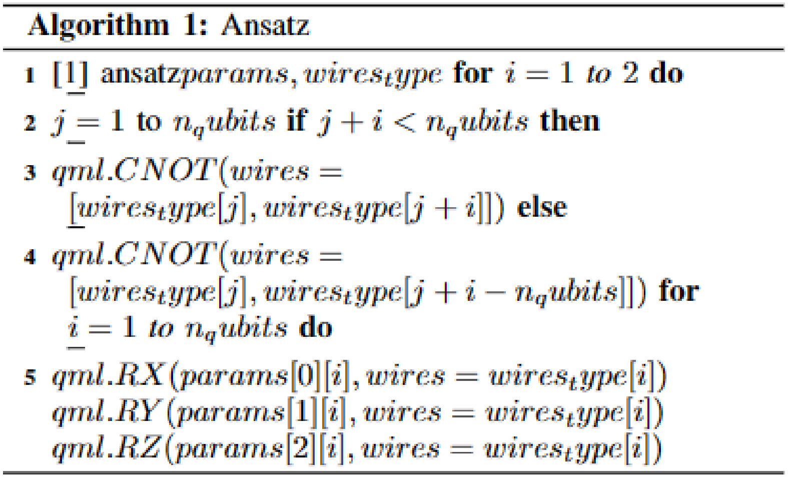


Figure 20. Architecture Of VQC

The classical data is fed through a Hadamard gate which preserves normalization and turns the data that it gets into a super positioned state. As we progress through the gates, there are rotational gates which rotate the phase of the qubit with respect to the y and z axis which manipulates the quantum bits to be suited and prepped for the variational quantum circuit. VQC acts as a hidden state which can be stacked together and are now processed using Multi-bit CNOT gates which manipulate the qubits based on the state of the controlled bit. This block has the ability to listen to parameters and perform operations at every time step and improve much like a traditional network does during backpropagation.

The variational quantum circuit however does not use the quantum state distances because this is a hybrid system. During back propagation a normal MSE loss function is used instead however, the specific system can be tuned into a task wherein it uses the fidelity distance. The system then converts the output into a measurement system which calculates the probability using the observable operator to calculate the probability with which the state collapses into a value to use it during classical processing.

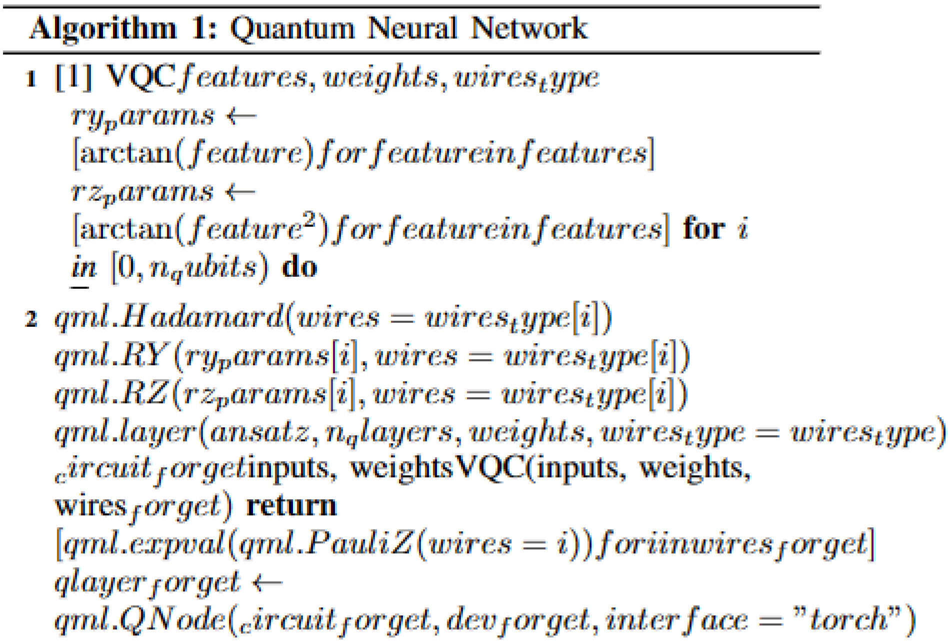


Algorithm 1: Ansatz

```
[1] ansatz$params, wires_type$ for $i = 1$ $to$ $2$ do
$j = 1$ to $n_qubits$ if $j + i < n_qubits$ then
$qml.CNOT(wires = [wires_type[j], wires_type[j + i]])$ else
$qml.CNOT(wires = [wires_type[j], wires_type[j + i - n_qubits]])$ for $i = 1$ $to$ $n_qubits$ do
$qml.RX(params[0][i], wires = wires_type[i])$
$qml.RY(params[1][i], wires = wires_type[i])$
$qml.RZ(params[2][i], wires = wires_type[i])$
```

Algorithm of the Ansatz

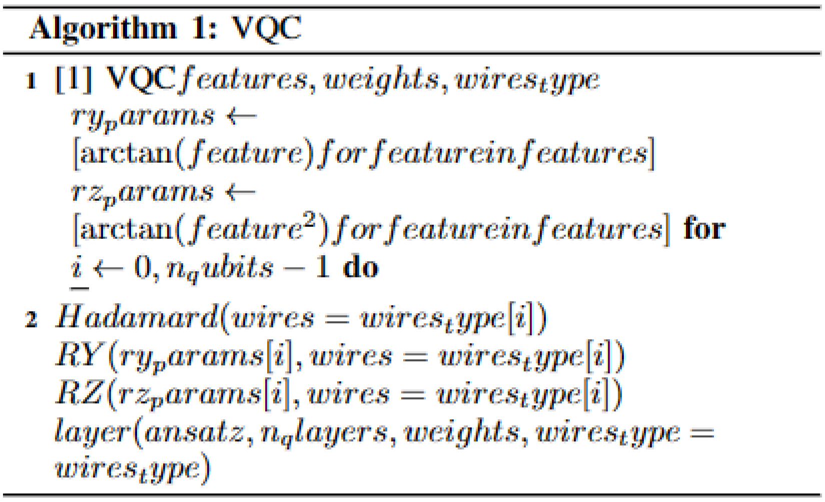
The following algorithms describe the inputs and the flow of the VQC as well as one of the circuit gates of the ansatz to understand how the probability measurement is calculated.



```
Algorithm 1: Quantum Neural Network
[1] VQC$features, weights, wires_{t}ype$
    $ry_{p}arams \leftarrow$
    [arctan($feature$)$forfeatureinfeatures$]
    $rz_{p}arams \leftarrow$
    [arctan($feature^2$)$forfeatureinfeatures$] for $i$
    in $[0, n_{q}ubits)$ do
        $qml.Hadamard(wires = wires_{t}ype[i])$
        $qml.RY(ry_{p}arams[i], wires = wires_{t}ype[i])$
        $qml.RZ(rz_{p}arams[i], wires = wires_{t}ype[i])$
        $qml.layer(ansatz, n_{q}layers, weights, wires_{t}ype = wires_{t}ype)$
        $_{c}ircuit_{f}orget$inputs, weightsVQC(inputs, weights,
        wires$_{f}orget$) return
        $[qml.expval(qml.PauliZ(wires = i))foriinwires_{f}orget]$
        $qlayer_{f}orget \leftarrow$
        $qml.QNode(_{c}ircuit_{f}orget, dev_{f}orget, interface = "torch")$
```

Algorithm of the quantum gate

As noticed, the function calculates the measurement value using the qml library's inbuilt expval method which calculates the probability with which the state collapses into a definite value.



```
Algorithm 1: VQC
[1] VQC$features, weights, wires_{t}ype$
    $ry_{p}arams \leftarrow$
    [arctan($feature$)$forfeatureinfeatures$]
    $rz_{p}arams \leftarrow$
    [arctan($feature^2$)$forfeatureinfeatures$] for
    $i \leftarrow 0, n_{q}ubits - 1$ do
        $Hadamard(wires = wires_{t}ype[i])$
        $RY(ry_{p}arams[i], wires = wires_{t}ype[i])$
        $RZ(rz_{p}arams[i], wires = wires_{t}ype[i])$
        $layer(ansatz, n_{q}layers, weights, wires_{t}ype =$
        $wires_{t}ype)$
```

Figure 23. Algorithm of VQC

REFERENCES

[1.] S. G. Steven Elsworth, “Time series forecasting using lstm networks: A symbolic approach,” 2020

[2.] J. S. Sepp Hochreiter, “Long short term memory,” Neural computation, 1997

[3.] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” Physica D: Nonlinear Phenomena, vol. 404, p. 132306, mar 2020

[4.] C. Fjellstr¨om, “Long short-term memory neural network for financial time series,” 2022.

[5.] D. Hopp, “Economic nowcasting with long short-term memory artificial neural networks (lstm),” 2022.

[6.] A. Srivastava and A. S, “Weather prediction using lstm neural networks,” in 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022, pp. 1–4.

[7.] A. S. N. Sima Siami-Namini, Neda Tavakoli, “A comparison of arima and lstm in forecasting time series,” 2018.

[8.] J. C.-B. David Peral Garc´ıaa and F. J. Garc´ıa-Penalvoc, “Systematic ˜ literature review: Quantum machine learning and its applications,” 2022

[9.] A. R.-K. TARIQ M. KHAN, “Machine learning: Quantum vs classical,” 2020.

[10.] H. Mizuno, “Research and development of annealing and gate-based quantum computing,” Development of Disruptive Technologies for Pioneering Next Generation, 2021.

[11.] M. A. Nielsen, “Quantum computation and quantum information,” Cambridge University Press, 2021.

[12.] B. R. Kazem, “The effect of pauli gates on the superposition for four-qubit in bloch sphere,” journal of kerbala university, 2020.

[13.] S. Mangini, “Quantum computing models for artificial neural networks,” EPL, 2021

[14.] R. D. Sipio, “The dawn of quantum natural language processing,” 2021.

[15.] M. Lin, “Optimal transformer modeling by space embedding for ionospheric total electron content prediction,” IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, 2022

[16.] K. Bee, “Quantum neural networks,” 2022

[17.] S. Shekhar, “Language identification framework in code-mixed social media text based on quantum lstm,” 2020.

[18.] J. Bausch, “Recurrent quantum neural networks,” 2020

[19.] Y.-L. L. F. Samuel Yen-Chi Chen, Shinjae Yoo, “Quantum long shortterm memory,” 2022.

[20.] P. Singh, “Fqtsfm: A fuzzy-quantum time series forecasting model,” 2021.

[21.] T. Y. Xiangyi Meng, “Entanglement-structured lstm boosts chaotic time series forecasting,” 2021

[22.] S. D. Dimitrios Emmanoulopoulos, “Quantum machine learning in finance: Time series forecasting,” 2022.

[23.] Nilesh Barde, D. T. (2011). Consequences and Limitations of Conventional Computers and their . *Leonardo Electronic Journal of Practices and Technologies*.

[24.] Jazaeri, F., Beckers, A., Tajalli, A., & Sallese, J. M. (2019, June). A review on quantum computing: From qubits to front-end electronics and cryogenic MOSFET physics. In 2019 MIXDES-26th International Conference" Mixed Design of Integrated Circuits and Systems" (pp. 15-25). IEEE.

[25.] Ku, J., Xu, X., Brink, M., McKay, D. C., Hertzberg, J. B., Ansari, M. H., & Plourde, B. L. T. (2020). Suppression of Unwanted Z Z Interactions in a Hybrid Two-Qubit System. _Physical review letters_, _125_(20), 200504.

[26.] Ferrando-Soria, J., Moreno Pineda, E., Chiesa, A., Fernandez, A., Magee, S. A., Carretta, S., ... & Winpenny, R. E. (2016). A modular design of molecular qubits to implement universal quantum gates. _Nature communications_, _7_(1), 11377.

[27.] Kiani, B. T., De Palma, G., Marvian, M., Liu, Z. W., & Lloyd, S. (2022). Learning quantum data with the quantum earth mover’s distance. _Quantum Science and Technology_, _7_(4), 045002.

[28.] Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). Quantum circuit learning. _Physical Review A_, _98_(3), 032309.

[29.] Piroli, L., Bertini, B., Cirac, J. I., & Prosen, T. (2020). Exact dynamics in dual-unitary quantum circuits. Physical Review B, 101(9), 094304.

[30.] Benedetti, M., Lloyd, E., Sack, S., & Fiorentini, M. (2019). Parameterized quantum circuits as machine learning models. _Quantum Science and Technology_, _4_(4), 043001.

[31.] Cho, S. (2022). Algorithms of quantum computation. In Quantum Computation and Quantum Information Simulation using Python: A gentle introduction. IOP Publishing.

[32.] Baaquie, B. E., & Kwek, L. C. (2023). Deutsch Algorithm. In _Quantum Computers: Theory and Algorithms_ (pp. 173-178). Singapore: Springer Nature Singapore.

[33.] Voichick, F., Li, L., Rand, R., & Hicks, M. (2023). Qunity: A Unified Language for Quantum and Classical Computing. Proceedings of the ACM on Programming Languages, 7(POPL), 921-951.

[34.] Chen, S. Y. C., Yoo, S., & Fang, Y. L. L. (2022, May). Quantum long short-term memory. In _ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)_ (pp. 8622-8626). IEEE.

[35.] Niu, X., Hou, Y., & Wang, P. (2017). Bi-directional LSTM with quantum attention mechanism for sentence modeling. In _Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24_ (pp. 178-188). Springer International Publishing.

[36.] A. Padha and A. Sahoo, "A Parametrized Quantum LSTM Model for Continuous Stress Monitoring," *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2022, pp. 261-266, doi: 10.23919/INDIACom54597.2022.9763118.

[37.] M. A. R. Suleman and S. Shridevi, "Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model," in *IEEE Access*, vol. 10, pp. 82456-82468, 2022, doi: 10.1109/ACCESS.2022.3196381.

[38.] Hao X, Liu Y, Pei L, Li W, Du Y. Atmospheric Temperature Prediction Based on a BiLSTM-Attention Model. *Symmetry*. 2022; 14(11):2470. https://doi.org/10.3390/sym14112470

[39.] Zhang, Yazhou & Song, Dawei & Li, Xiang & Zhang, Peng & Wang, Panpan & Rong, Lu & Yu, Guangliang & Wang, Bo. (2020). A Quantum-like Multimodal Network Framework for Modeling Interaction Dynamics in Multiparty Conversational Sentiment Analysis. Information Fusion. 62. 10.1016/j.inffus.2020.04.003.

[40.] X. Wang, X. Wang and S. Zhang, "Adverse Drug Reaction Detection From Social Media Based on Quantum Bi-LSTM With Attention," in IEEE Access, vol. 11, pp. 16194-16202, 2023, doi: 10.1109/ACCESS.2022.3151900.

[41.] Z. Karevan and J. A. K. Suykens, "Spatio-temporal stacked LSTM for temperature prediction in weather forecasting," 2018, arXiv:1811.06341.

[42.] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015.

[43.] M. Akram and C. El, "Sequence to sequence weather forecasting with long short-term memory recurrent neural networks," Int. J. Comput. Appl., vol. 143, no. 11, pp. 7–11, 2016.

[44.] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, arXiv:1406.1078.