Jurnal
Teknologi

Full Paper

QUANTIZED LSTM FOR PREDICTIVE THERMAL MONITORING OF TEMPERATURE OF LITHIUM-ION BATTERY

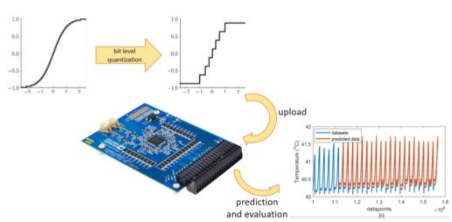
Kian Kok David Hong[a], Mohd Rashidi Salim[a*], Joshu Leonardy[a], Nur Aisyah Athirah Mohd Zali[a], Farabi Iqbal[a], Mohd Haniff Ibrahim[a], Hadi Manap[b]

[a]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan Campus, 26600 Pekan, Pahang, Malaysia

Article history
Received
24 September 2024
Received in revised form
6 April 2025
Accepted
13 May 2025
Published Online
27 February 2026

*Corresponding author
mrashidi@utm.my

Graphical abstract



Abstract

The large memory requirements and low inference speed renders the deployment of Long short-term memory (LSTM) impractical for predictive monitoring in battery thermal management system. Thus, this study proposes quantization aware training, where the weights and biases were quantized using ternary quantization scheme and while the activation function were quantized using 8-, 6- and 4-bit level schemes. The various quantization strategies were evaluated with statistical analysis such as mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) and train and validated with oxford battery dataset. In addition, the various bit level schemes were compared in terms of performance.As 4-bit level level offers lowest memory footprint. This space saving is done where 24-bit = 16 values are assigned for the activation unit and unable to cover the whole spectrum compared to full precision, thus results in large mean accuracy loss. On the other hand, 6-bit level quantization proves to be better performance with relatively smaller memory footprint compared to 8-bit level quantization.Thus, 6-bit level quantization is suitable for low power edge application for battery management system.

Keywords: Lithium-ion battery, temperature, long short-term memory, quantization

Abstrak

Storan data yang besar dan kelajuan process jangkan yang lambat menjadikan penggunaan Long Short Term Memory (LSTM) tidak praktikal untuk dalam sistem pengurusan bateri. Oleh itu, kajian ini mencadangkan kaedah Quantization Aware training, di mana kuantiti-kuantiti weight dan bias dikuantisasi menggunakan skema pengkuantitian ternary dan manakala fungsi activation dikuantisasi menggunakan skema tahap 8-, 6- dan 4-bit. kaedah-kaedah pengkuantitian ini dinilai dengan ralat min kuasa dua (MSE), ralat mutlak min (MAE) dan ralat peratusan min mutlak (MAPE) dan disahkan dengan dataset bateri oxford sebagai dataset penilaian. Di samping itu, pelbagai skema tahap bit telah dibandingkan dari segi prestasi. Memandangkan tahap 4-bit menawarkan storan data yang lebih rendah. Penjimatan ruang ini dilakukan di mana 24-bit = 16 nilai ditetapkan untuk kuantiti activation unit dan tidak dapat menampung keseluruhan spektrum berbanding dengan ketepatan penuh, dengan itu mengakibatkan kehilangan ketepatan min. Sebaliknya, pengkuantitian tahap 6-bit terbukti sebagai prestasi yang lebih baik dengan jejak memori yang agak kecil berbanding dengan pengkuantitian tahap 8-bit. Oleh itu, pengkuantitian tahap 6-bit adalah sesuai untuk aplikasi kelebihan kuasa rendah untuk sistem pengurusan bateri.

Kata kunci: Aliran trafik, pesawat tanpa pemandu, *quadrotor,* video masa nyata

1.0 INTRODUCTION

In order to meet the incessant increasing adoption of lithium-ion battery due to the demand from society to move into a greener and more sustainable source of energy, this has spur innovation in lithium-ion battery technology that has high energy density with lower overall weight. The candidate for the future of rechargeable battery technology, especially the anode is lithium metal [1, 2, 3, 4] such as lithium air and lithium sulphur battery as shown in Figure 1.

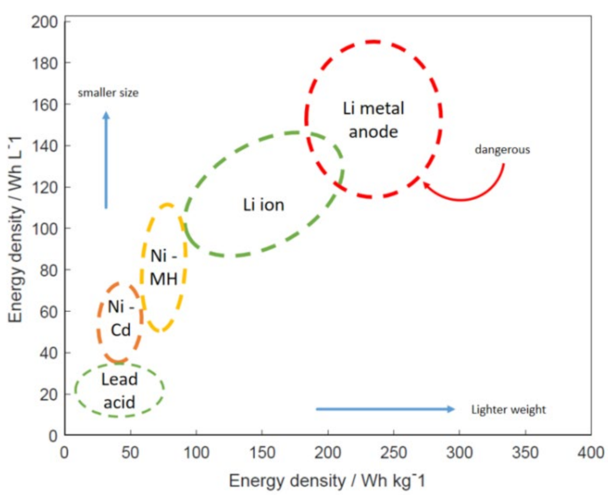


Figure 1 Development trajectory of rechargeable battery technology based on energy density

Nevertheless, such prospects also raise concern especially safety issue. The material anode of lithium metal battery used is lithium metal, unlike conventional lithium-ion battery, where the material used are graphite. Lithium metal as anode is found to be more aggressive [1, 5] during the charging-discharging process and generate excessive heating. Thus, this hasten the decomposition of electrolyte and separator, and finally results in thermal runaway [6][7]. Substantial effort in the development of electrolyte, separator and cooling technology to mitigate this issue have been proposed [8, 9]. Nonetheless, the occasion external mechanical, electrical and thermal abuses cause the thermal behavior to be unpredictable. Thus, this necessitates the development of anomaly prediction and early forecasting of battery temperature to predict its eventual temperature [10], [11], [12], [13]. Deploying such predictive scheme raise the safety level standard of battery usage and unwanted dangerous and hazard occurrence such as explosion can be avoided.

There are many predictive scheme have been proposed, developed and reviewed [29, 30, 31]. Recently, data driven method is gaining increasing attention as complexity of batteries' physics and chemistry is not required and the problem itself ultimately is data science. Since prediction is pertaining to time series forecasting, Recurrent neural network (RNN) which is a deep learning methods such as Long Short Term Memory (LSTM) have demonstrate excellent performance in prediction of battery states such as state of health degradation [14, 15, 16, 17] and external heating of lithium-ion battery [18, 19, 20]. These battery states can be modeled as sequential data especially charging-discharging cycle that reaches more than 200 cycles, which makes LSTM suitable for the task as the internal configuration of LSTM are ideal to due to memory and gate mechanism, which helps in capture features for long-term dependencies.

The basic features that is the building block of neural network is the weight, bias and activation unit parameter, which relates how important is input to the neurons, bias helps the neuron to recognize pattern that not always appear often or can be considered as unpredictable. These 3 parameters (weight, bias and activation) often expressed in a single equation as shown in Figure 2.

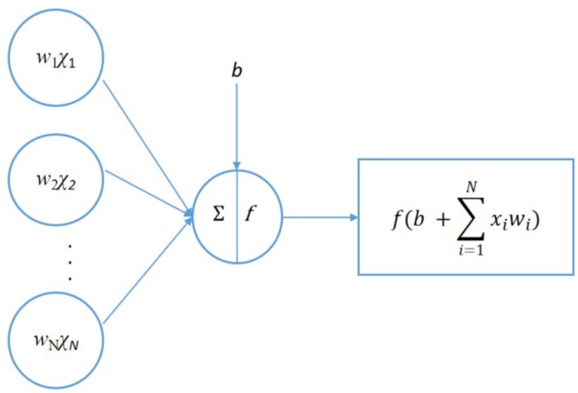


Figure 2 The mathematical expression of neural network where *w* is weight, *χ* is input, *b* is the bias and *f* is the activation unit

The values of weight, bias and activation depend on the training datasets. On the other hand, the building block of LSTM unit consist of 4 trainable parameters- input gate, output gate, forget gate and input modulation gate. Each gate consists of unique weight and bias. The ever changing temperature pattern due to degradation, raise the number of LSTM units required for LSTM model to effectively learn it. This is evidenced by the public datasets available [Inevitably, the incurred memory requirement of LSTM model is very large and may be unsuitable to be deployed for edge application.

One way to improve the inference speed of LSTM, especially if the hardware platform has a lower memory storage resource is to adopt quantization technique. Quantization of LSTM model can improve inference speed and leads to lower memory requirement, which is suitable for embedded or edge application. This is done by transforming the weights, biases and activation of a LSTM mode from full precision, such as 32-bit floating point to a lower bit scale level such as 8-bit integer or lower. The reported results pertaining to quantization of RNN-LSTM model are mostly in the biomedical application [21, 22, 23].

Inspired by the biomedical application, this paper aims to develop and evaluate the accuracy of quantized LSTM model for long term temperature forecasting in lithium-ion battery. The contribution of this paper are as follow.

1. To develop quantized LSTM model for long term thermal forecasting by using oxford battery datasets .
2. To compare and evaluate the quantized network with the full precision network in terms of mean error, training and testing loss.
3. To elucidate the various quantized strategy and its impacts on the datasets.

2.0 METHODOLOGY

The basic building block of LSTM, which is a hidden unit cell are inputs, weights, biases and activation. Only the weights and biases are the learnable parameters during the training process. Unlike artifiicial neural network, it has three input parameters (h_{t-1}, c_{t-1}, x_t). and three output parameters. (h_t, c_t, o_t). Such extra input and output parameters are attributed to gates that responsible for control, addition and removal of information in the cell states of LSTM.

As shown in Figure 3, there are three main gates that control the operation of LSTM such as forgetting gate, where its function is to forget the information from the previous internal cell, C_{t-1} based on the previous hidden state, h_{t-1} and the current input value, x_t.

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{1}$$

The input gate decides the incoming values that needs to be added into the cell state based on the value of sigmoid, σ. The incoming values are previous hidden state, h_{t-1} and current input value x_t.

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{2}$$

In addition, a new candidate vector, $\tilde{c}_t$, is created.

$$\tilde{c}_t = tanh(W h_{t-1} + U x_t + b) \tag{3}$$

The cell state, c_t is then updated by forgetting the previous cell state, c_{t-1} , and adding the new memory using the newly created candidate vector, $\tilde{c}_t$.

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{4}$$

The output gate o_t determines the next hidden state value that contains the previous information.

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{5}$$

The cell output h_t , which is passed into the next time step as an input.

$$h_t = o_t \circ tanh(c_t) \tag{6}$$

Where $\boldsymbol{W_f}$, $\boldsymbol{W_i}$ and $\boldsymbol{W_o}$ are the weight values of forget, input and output respectively. While $\boldsymbol{b_f}$, $\boldsymbol{b_i}$, and $\boldsymbol{b_o}$ are the biases for forget, input and output respectively and $\boldsymbol{\sigma}$ is the activation function.

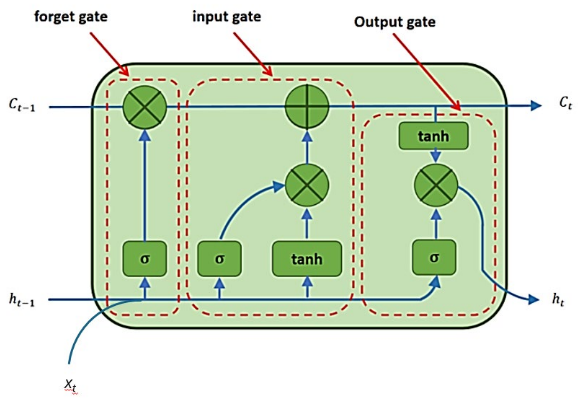


Figure 3 The basic cell of LSTM with forget, input and output gates

The LSTM model are quantized. Although the weight of LSTM layer with full precision such as 32-bit floating point, has higher accuracy and low training loss, it requires large memory and is not ideal for embedded application. In order for the LSTM model to be practical enough for embedded application, quantization of weights, biases and activation are necessary. Thus, this can reduce the network latency and lower the power consumption.

In this project, we adopt the ternary quantization scheme for the weight and biases. According to the framework of LSTM, for each input, output and forget gates require independent weights and biases. Thus, this justifies the use of ternary quantization scheme to trim the weight and bias parameters in effort to produce small memory footprint LSTM model.

However, weights and biases that have been quantized suffers from accuracy loss and large overall mean error compared to LSTM model with full precision. Thus, the activation function of LSTM models was not subjected to the same quantization scheme as weight and bias, instead data representation in integer format such as 8, 6 and 4 bits were adopted for activation function to avoid further accuracy loss. As activation functions are shared for the input, output and forget gate, this do not impact the overall memory footprint.

Ternary quantization scheme is considered a framework of quantization aware training (QAT). It was introduced by Li & Liu (2016) [24] where the training of LSTM model is with the 3 quantized values of either -1, 0 and 1. These quantized values are determined by quantization threshold as shown in the expression below. Ternary quantization weight minimize the accuracy loss by adopting 3 quantized values.

$$Q(\omega) = \begin{cases} -1: & \omega \leq -\Delta \\ 0: & |\omega| < \Delta \\ 1: & \omega > \Delta \end{cases}$$

The quantization threshold, ±Δ is to minimize the mean distance between full precision and ternary quantized weight.

For the activation function in the LSTM layers, applying ternary quantization causes large loss to the weight and biases [12]. Thus, quantization with fixed points data representation were adopted with 8-, 6-, and 4-bits integers. Such a strategy is commonly known as mixed precision quantization strategy, where certain components such as activation function, have different bit quantization.

The LSTM network framework is summarized in Figure 4. The four steps are widely adopted in many LSTM network development, which consists of data preprocessing of temperature historical profile of charging and discharging of lithium battery, feature extraction to capture local fluctuations, model training with full precision and half-precision (quantized), and validation stage with statistical analysis.

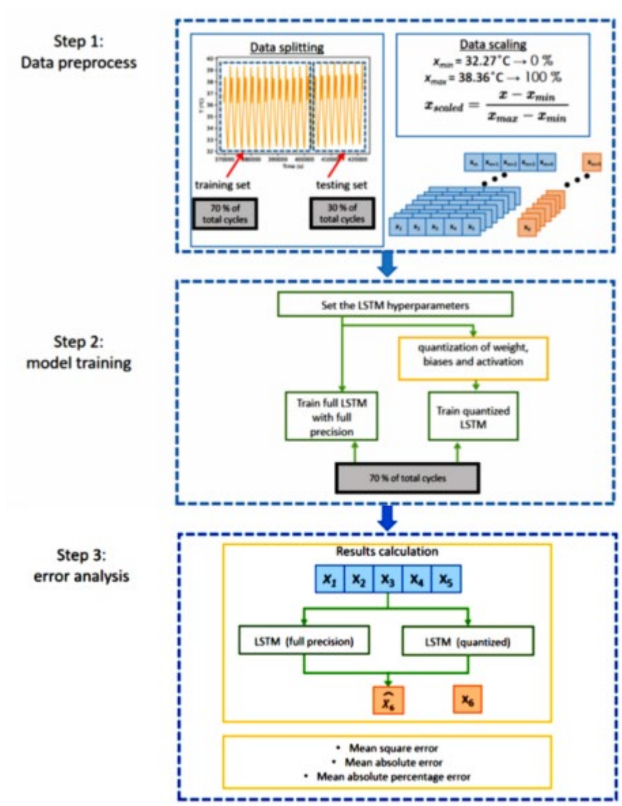


Figure 4 Flow chart of the development of LSTM with full precision and quantized weights, biases and activation

Python programming language with TensorFlow library was used to develop the LSTM network. It consists of 2 layers as shown in Figure 5.

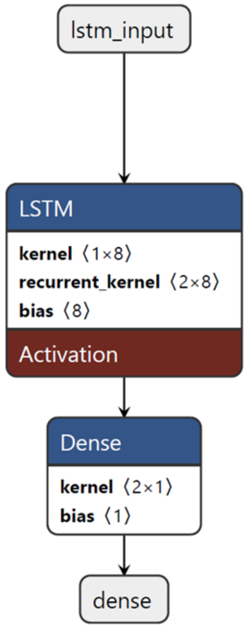


Figure 5 the architecture of LSTM model with input size of 5x1, LSTM layer size of 32x1 and dense layer size of 8x1

In this work, the datasets used were the temperature profile of long-term battery ageing tests (charging and discharging for 77 cycles) of lithium pouch cell with capacity of 740 mAh from Kokam manufacturer (SLPB533459H4). The discharging ratings for the battery ageing tests were 1C. The environment of charging and discharging was at 40 °C. The provided public domain datasets [25] were provided by oxford researchers and is widely utilized to benchmark the performance of machine learning model. Table 1 below summarizes the characteristics of datasets.

Table 1 Information of Oxford battery degradation datasets

Category	Details
Battery rated capacity	740mAh
Charging and discharging rating	1C (740mA)
Training dataset labels	Cell 1
Validation dataset labels	Cell 2, Cell 3, Cell 4, Cell 5, Cell 6 and Cell 7

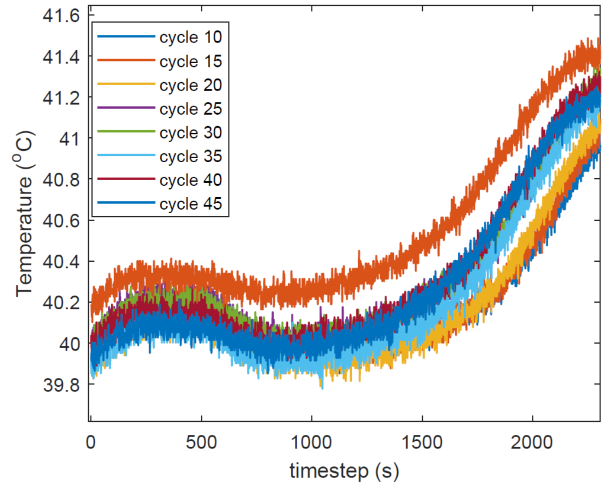


Figure 6 Selected temperature datasets during 1C discharging rate for Cell 1 for purpose of training and testing of LSTM full precision bits and quantized bits

Local fluctuations of battery temperature profile is inevitable and highly irregular as battery went through periodic of charging and discharging process as shown in Figure 6. Thus, it is essential to adopt sliding window method to minimize the overall mean error, where the local fluctuations of data trend can be learned and its features can be extracted by the LSTM model other than the overall trend. This was accomplished where the input array was iterated over a sequence of data. This involves the dividing of data into overlapping windowes of a fixed size, and processing each window independently.

3.0 RESULTS AND DISCUSSION

The accuracy of quantized LSTM is evaluated by the following statistical metrics. Three evaluation metrics are adopted to evaluate the accuracy of LSTM model.

Mean absolute Error is to quantify the errors between the values provided by datasets and predicted value. The purpose of Mean absolute percentage error (MAPE) is to gauge the relative error between the predicted temperature value and the actual temperature in percentage form . Finally, mean square error measures the average squared difference between the predicted temperature and temperature from the datasets. All of the statistical metrics are as follows.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{T}_i - T_i\right|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{T}_i - T_i}{T_i}\right| \times 100\%$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(T_i - \widehat{T}_i\right)$$

In order for the quantized LSTM model to achieve acceptable accuracy, the MAE, MAPE and MSE should be as low as possible.

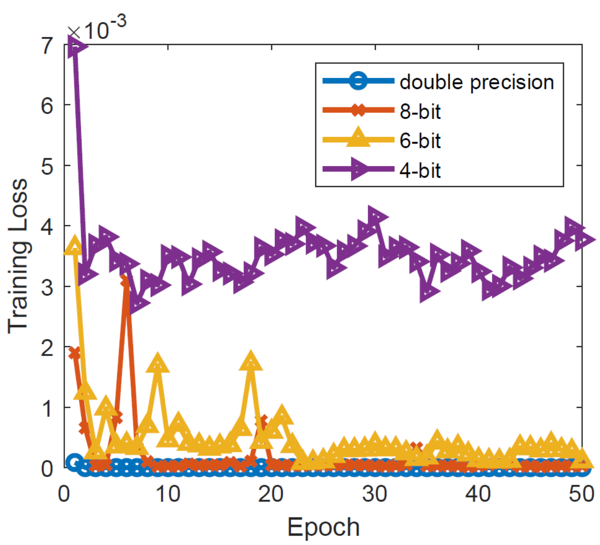
Before the experiment was conducted, the dataset was split in different time steps for the input and output training data. For the input, five time steps are constructed as a set of sequences for input and one time step are constructed as a set of sequences for output.

$$x_1 = [d_1, d_2, d_3, d_4, d_5]\,, y_1 = [d_6]$$
$$x_2 = [d_2, d_3, d_4, d_5, d_6]\,, y_2 = [d_7]$$
$$\vdots$$
$$x_N = [d_N, d_{N+1}, d_{N+2}, d_{N+3}, d_{N+4}]\,, y_N = [d_{N+5}]$$

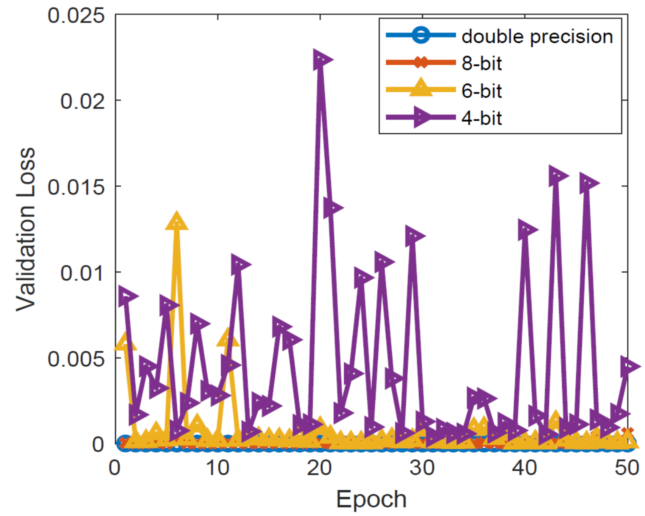
Although it is possible for the sliding window fixed length to continue to increment, nevertheless It is important to keep the number of training epochs at an acceptable range to prevent the LSTM network from over fitting. It is essential to monitor the training loss to avoid the network from diverging or showing increasing loss.

Besides, data normalization was done to reduce the overall training loss during the data preprocessing stage.

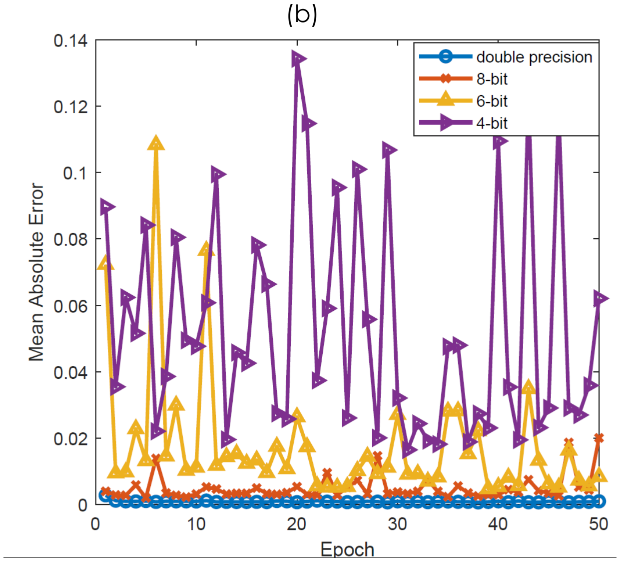
As shown in Figures 7(a) and 7(b), the training and validation loss curve for quantization bits shows improvement as the bit level increase from 4 bit to 8 bits. Despite the large truncation of data bit for weight, biases and activation, the performance of quantized network 6-bit quantization shows adequate performance in predicting the long-term trend of battery temperature. Figures 7(c), (d), and (e) show the MAE, MSE, and MAPE curves, respectively, for the full-precision and quantized (8-, 6-, and 4-bit) LSTM models using a batch size of 5 and 50 epochs.



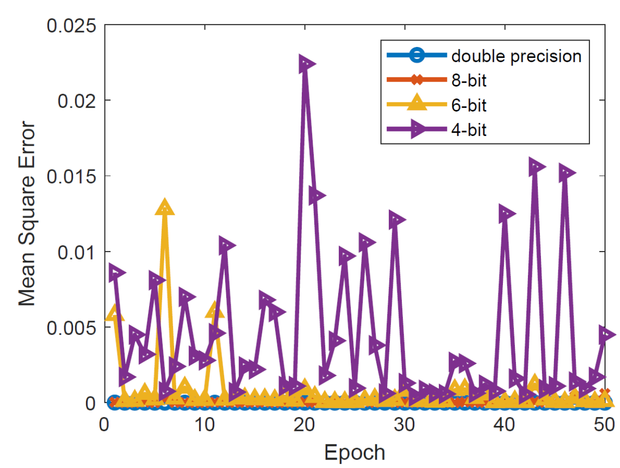
(a)



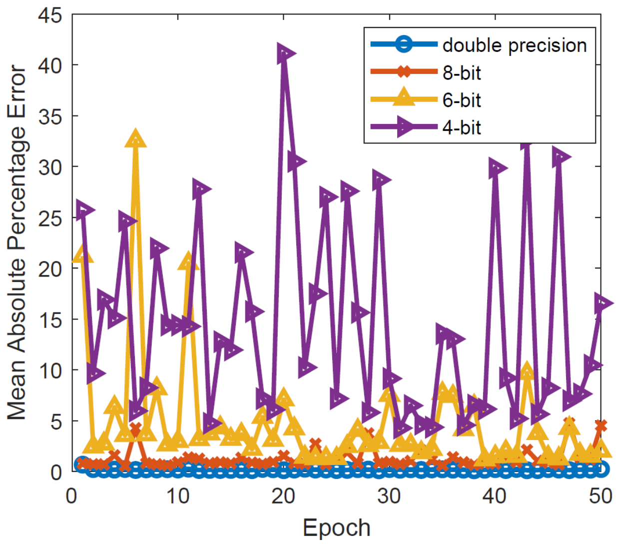
(b)



(c)

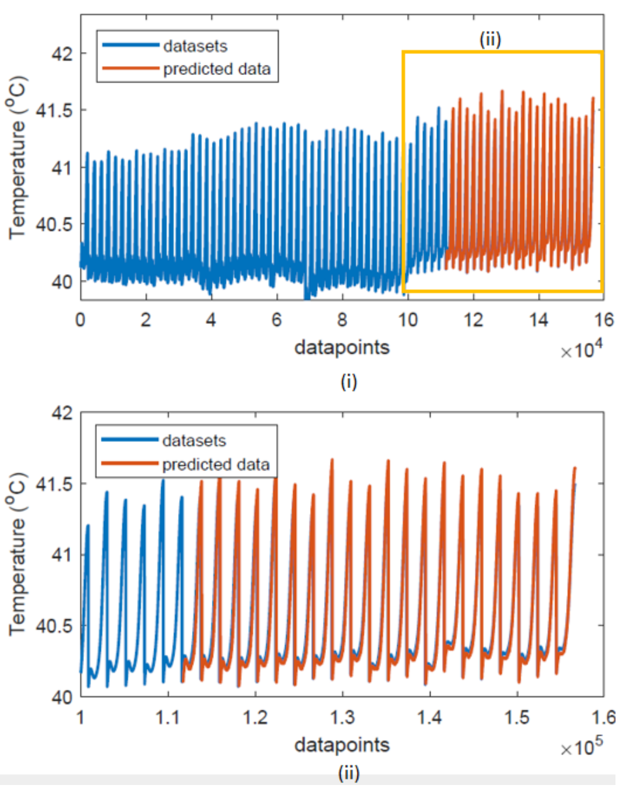


(d)

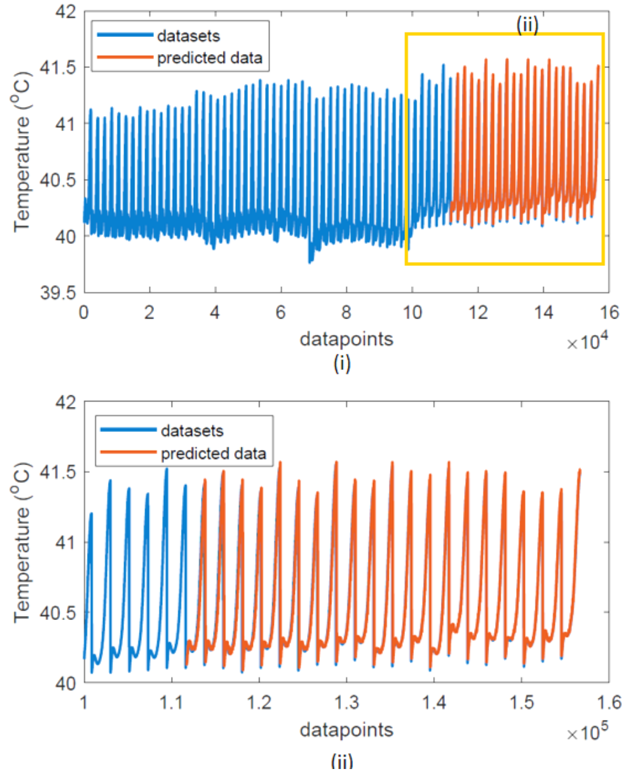


(e)

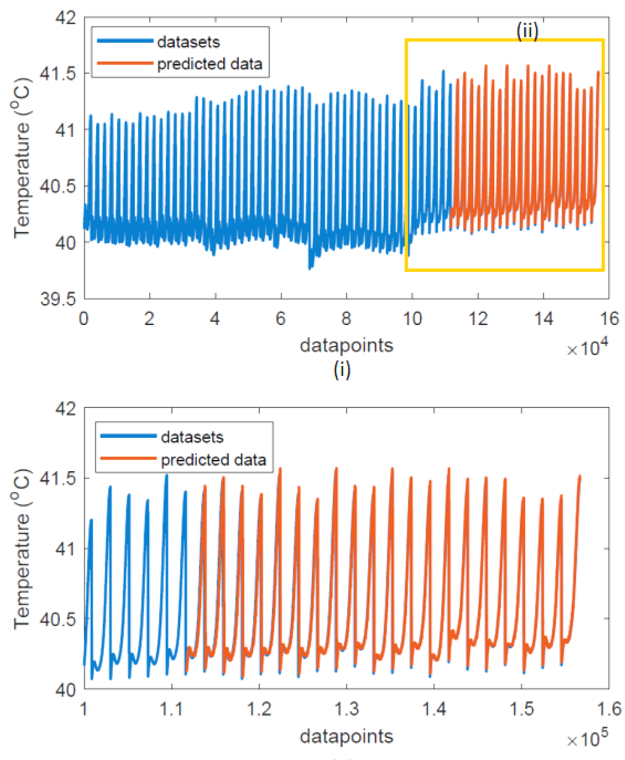
Figure 7 (a)Training loss, (b) validation loss (c) mean absolute error, (d) mean square error and (e) mean absolute percentage error curves for full precision, 8-bit, 6-bit, 4-bit quantization for activation and dense layer LSTM with epochs of 50 and batch size of 5



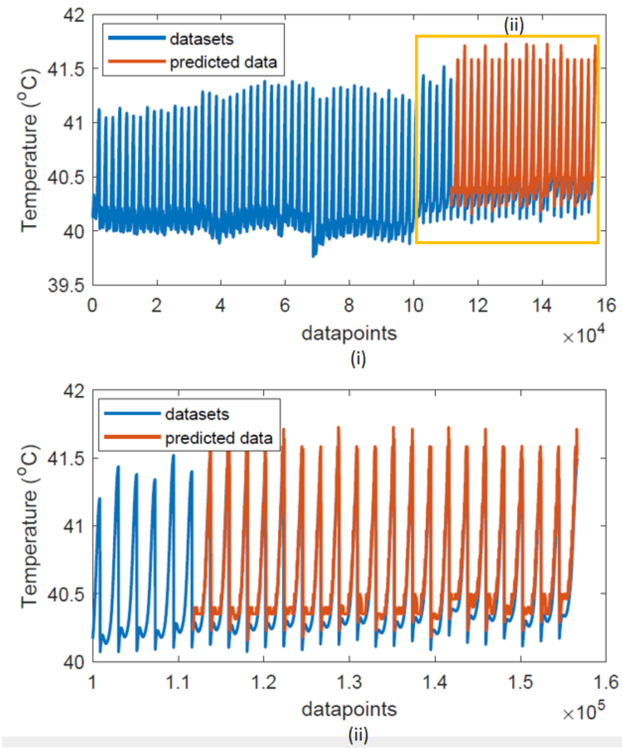
(a)



(b)



(c)



(d)

Figure 8 Evaluation of LSTM model with (a) full precision (b) 8-bit (c) 6-bit (d) 4-bit activation and dense layer

In this subsection, the various quantization bit level framework and full precision. In addition, to ensure the model is reliable, it is essential to verify its accuracy through the inference of datasets that the model were not trained. Based on Figure 7, it is obvious that the deviation of predicted data increases as the quantization level decreases. Although the bit level quantization decreases linearly, the MSE, MAE and MAPE shows slow increment.

The 4-bit level quantization scheme shows significant fluctuations as shown in Figure 8(d). Such fluctuations are due to activation oscillation [27], which can lead to poor validation accuracy. According to M. Nagel *et al*. (2022), as lower bit-width quantization scheme results in larger interval between the quantization levels. As shown in Figure 9, activation units are distributed evenly especially values ranging from -0.05 to 0.05. the likelyhood of mean error is higher when the quantized bit level is lower . The consequence of quantized bit level of activation unit is some portion of activation unit values are replaced with the adjacent threshold activation unit value that has been assigned by quantization framework.

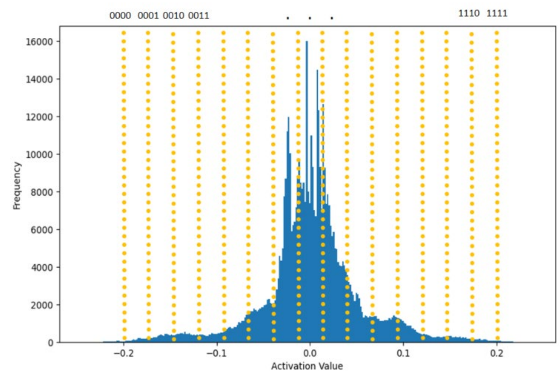
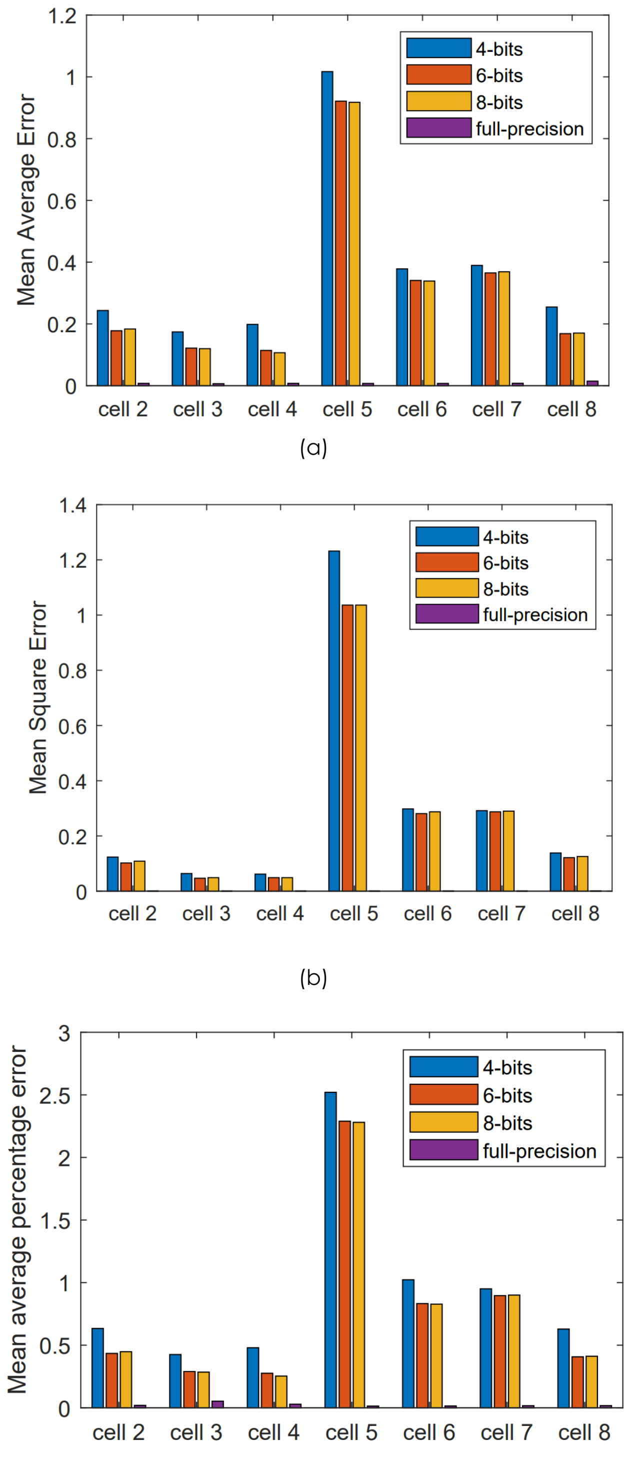


Figure 9 the the distribution of full precision activation unit of LSTM model and the expected 4-bit level quantization treshold value. When bit level quantization is applied to the activation unit during quantization aware training,not all of the activation units in full precision are assigned to the quantization level values

The oscillation of activation due to 4-bit level quantization induces larger mean percentage error approximately 10 % on average and shows no improvement despite larger training epoch. Such large variation of output prediction unfit for the development battery temperature tracking and early warning system. As for 6-bit level quantization, it did not suffers on average from large oscillation with total of 50 training epochs as shown in Figure 8(c). This makes the 6-bit level quantization option attractive for the predictive monitoring of battery temperature in embedded application. As for 8-bit, the average MSE is, which is almost close to 32-bit floating point although there is some oscillations as shown in Figure 8(b). For full precision of activation and dense layer as shown in Figure 8(a), naturally it does not suffer from any oscillation, but large memory footprint is needed to support such high performance of LSTM model.

This section concerns the robustness of the proposed LSTM network by testing it with datasets that have not been trained. As thermal stability of battery tends to drift especially after long series of cycling, this necessitates the testing of LSTM network with diverse set of datasets.

In this work, the testing of cell 2, 3, 4, 5, 6 and 7 of all of the datasets[28] are only for the quantized LSTM network.



(a)

(b)

(c)

Figure 10 Performance metrics of (a)MAE (b) MSE and (c)MAPE of 4-,6- and 8-bit quantization LSTM model using data of cell 2,3,4, 5, 6 and 7 as evaluation datasets

Table 2 MAE metrics for 4-,6- and 8-bit quantization LSTM model using data of cell 2,3,4, 5, 6 and 7 as evaluation datasets

Category	4-bits	6-bits	8-bits	Full-precision
Cell 2	0.243446	0.177903	0.183521	0.00770
Cell 3	0.174157	0.121723	0.11985	0.00659
Cell 4	0.198502	0.114232	0.106742	0.007668
Cell 5	1.016854	0.921348	0.917603	0.007380
Cell 6	0.378277	0.340824	0.338951	0.007420
Cell 7	0.389513	0.365169	0.368914	0.00790
Cell 8	0.254682	0.168539	0.170412	0.01473

Table 3 MAPE metrics for 4-,6- and 8-bit quantization LSTM model using data of cell 2,3,4, 5, 6 and 7 as evaluation datasets

Category	4-bits	6-bits	8-bits	Full-precision
Cell 2	0.63348163	0.43439	0.447964	0.01939
Cell 3	0.42534	0.2896	0.28507	0.0527
Cell 4	0.47964	0.27602	0.2534	0.02848
Cell 5	2.520362	2.2896	2.28054	0.01376
Cell 6	1.022624	0.8326	0.82805	0.0145
Cell 7	0.950226	0.89593	0.900452	0.0167
Cell 8	0.6289593	0.40724	0.411765	0.0172

Table 4 MSE metrics for 4-,6- and 8-bit quantization LSTM model using data of cell 2,3,4, 5, 6 and 7 as evaluation datasets

Category	4-bits	6-bits	8-bits	Full-precision
Cell 2	0.123404	0.10213	0.1085106	0.000103
Cell 3	0.06383	0.04681	0.0489361	0.0000743
Cell 4	0.0617	0.048936	0.04893617	0.0000986
Cell 5	1.231915	1.03617	1.03617	0.0000849
Cell 6	0.297872	0.28085	0.287234	0.0000884
Cell 7	0.29149	0.28723	0.289855	0.0000959
Cell 8	0.13830	0.12128	0.125532	0.000389

The MAE, MAPE, and MSE metrics for the 4-, 6-, and 8-bit quantized LSTM models are summarized in Tables 2, 3, and 4. These models were evaluated using data from cells 2, 3, 4, 5, 6, and 7. Overall, the quantization bit level scheme from 4 to 8 shows acceptable performance as reflected in the standard error metrics such as mean square error (MSE), mean average error (MAE) and mean average percentage error (MAPE). The approximate aver age value of MSE stood around 0.05 to 0.3 for all cells with the exception of cell 5, where three of the bit level schemes stood on average at 1.0. Likewise, MAE and MAPE stood approximately 0.10 ~ 0.40 and 0.25 % ~ 1.0 % respectively with the exception of cell 5 stood around 2.0 and 3.0 % respectively.

Based on Figure 10(a), (b) and (c), 4-bit quantization level scheme results in slightly higher MSE, MAE and MAPE compared to 6 and 8 bit quantization level. On the other hand, the MSE, MAE and MAPE metrics for 6- and 8-bit level quantization do not show noticeable difference. Thus, we can conclude that lowering the 6 bit quantization level on the activation has lowest model size with lowest MSE, MAE and MAPE metrics.

On the other hand for the full precision bit level, the performance of LSTM long term temperature prediction shows excellence MSE, MAE and MAPE metrics.

4.0 CONCLUSION

In this study, we developed quantized LSTM model with 4, 6 and 8-bit level quantization scheme. Its accuracy were tested with mean square error, mean abosolute percentage error and mean average error. It is revealed that the 4-bit quantization level scheme results in large oscillations during the training phase, which can lead to large erroneous prediction value as optimized training epoch is hard to determined. While 6 and 8 bits quantized LSTM model showed almost good accuracy in comparison with full precision LSTM model, 6-bits quantized LSTM model is suffice in terms of accuracy without incurring large memory footprint.

Acknowledgement

The author acknowledge the financial support received from Universiti Teknologi Malaysia under the UTM Encouragement Research Grant Scheme (Q.J130000.3823.42J20) and UTM Nexus Postgrad Grant Scheme (Q.J130000.5323.00L99).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Wang, R., Cui, W., Chu, F. and Wu, F. 2020. Lithium Metal Anodes: Present and Future. *Journal of Energy Chemistry*. 48: 145–159.

[2] Liu, D.H., Bai, Z., Li, M., Yu, A., Luo, D., Liu, W., Yang, L., Lu, J., Amine, K. and Chen, Z. 2020. Developing High Safety Li-metal Anodes for Future High-energy Li-metal Batteries: Strategies and Perspectives. *Chemical Society Reviews*. 49(15): 5407–5445.

[3] Chen, K., D. Y. Yang, G. Huang, and X. B. Zhang. 2021. Lithium–Air Batteries: Air-Electrochemistry and Anode Stabilization. *Accounts of Chemical Research*. 54(3): 632–641.

[4] Zhao, H., N. Deng, J. Yan, W. Kang, J. Ju, Y. Ruan, X. Wang, X. Zhuang, Q. Li, and B. Cheng. 2018. A Review on Anode for Lithium–Sulfur Batteries: Progress and Prospects. *Chemical Engineering Journal*. 347: 343–365.

[5] Piao, Z., H. R. Ren, G. Lu, K. Jia, J. Tan, X. Wu, Z. Zhuang, Z. Han, C. Li, R. Gao, and X. Tao. 2023. Stable Operation of Lithium Metal Batteries with Aggressive Cathode Chemistries at 4.9 V. *Angewandte Chemie International Edition* 62 (15): e202300966.

[6] Zhang, Q. K., X. Q. Zhang, H. Yuan, and J. Q. Huang. 2021. Thermally Stable and Nonflammable Electrolytes for Lithium Metal Batteries: Progress and Perspectives. *Small Science*. 1(10): 2100058.

[7] Zhang, S., Z. Shen, and Y. Lu. 2021. Research Progress of Thermal Runaway and Safety for Lithium Metal Batteries. *Acta Physico-Chimica Sinica*. 37: 2008065.

[8] Li, M., K. Wang, F. Shen, Y. Bai, C. Yang, J. Liu, J. Yan, Y. Chen, and X. Han. 2023. An Electrolyte-Locked Thermally Stable Composite Separator for Safe Lithium-Ion Batteries. *Journal of Alloys and Compounds*. 938: 168543.

[9] l'Abee, R., F. DaRosa, M. J. Armstrong, M. M. Hantel, and D. Mourzagh. 2017. High-Temperature Stable Li-Ion Battery Separators Based on Polyetherimides with Improved Electrolyte Compatibility. *Journal of Power Sources*. 345: 202–211.

[10] Zhu, S., C. He, N. Zhao, and J. Sha. 2021. Data-Driven Analysis on Thermal Effects and Temperature Changes of Lithium-Ion Battery. *Journal of Power Sources*. 482: 228983.

[11] Al Miaari, A., and H. M. Ali. 2023. Batteries Temperature Prediction and Thermal Management Using Machine Learning: An Overview. *Energy Reports*. 10: 2277–2305.

[12] Li, M., C. Dong, B. Xiong, Y. Mu, X. Yu, Q. Xiao, and H. Jia. 2022. STTEWS: A Sequential-Transformer Thermal Early Warning System for Lithium-Ion Battery Safety. *Applied Energy*. 328: 119965.

[13] Li, M., C. Dong, X. Yu, Q. Xiao, and H. Jia. 2021. Multi-Step Ahead Thermal Warning Network for Energy Storage System Based on the Core Temperature Detection. *Scientific Reports*. 11(1): 15332.

[14] Ma, Y., C. Shan, J. Gao, and H. Chen. 2022. A Novel Method for State of Health Estimation of Lithium-Ion Batteries Based on Improved LSTM and Health Indicators Extraction. *Energy*. 251: 123973.

[15] Van, C. N., and D. T. Quang. 2023. Estimation of State of Health and Internal Resistances of Lithium-Ion Battery Based on LSTM Network. *International Journal of Electrochemical Science*. 18(6): 100166.

[16] Yayan, U., A. T. Arslan, and H. Yucel. 2021. A Novel Method for State of Health Prediction of Batteries Based on Stacked LSTM with Quick Charge Data. *Applied Artificial Intelligence*. 35(6): 421–439.

[17] Kaur, K., A. Garg, X. Cui, S. Singh, and B. K. Panigrahi. 2021. Deep Learning Networks for Capacity Estimation for Monitoring State of Health of Li-Ion Batteries for Electric Vehicles. *International Journal of Energy Research*. 45(2): 3113–3128.

[18] Jiang, Y., Y. Yu, J. Huang, W. Cai, and J. Marco. 2021. Li-Ion Battery Temperature Estimation Based on Recurrent

Neural Networks. *Science China Technological Sciences*. 64(6): 1335–1344.

[19] Zhu, S., C. He, N. Zhao, and J. Sha. 2021. Data-Driven Analysis on Thermal Effects and Temperature Changes of Lithium-Ion Battery. *Journal of Power Sources*. 482: 228983.

[20] Yao, Q., D. D. C. Lu, and G. Lei. 2022. A Surface Temperature Estimation Method for Lithium-Ion Battery Using Enhanced GRU-RNN. *IEEE Transactions on Transportation Electrification.* 9(1): 1103–1112.

[21] Mirsalari, S. A., N. Nazari, S. A. Ansarmohammadi, S. Sinaei, M. E. Salehi, and M. Daneshtalab. 2021. ELC-ECG: Efficient LSTM Cell for ECG Classification Based on Quantized Architecture. In *Proceedings of the 2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–5.

[22] Rocha, L. G., D. Biswas, B. E. Verhoef, S. Bampi, C. Van Hoof, M. Konijnenburg, M. Verhelst, and N. Van Helleputte. 2020. Binary CorNET: Accelerator for HR Estimation from Wrist-PPG. *IEEE Transactions on Biomedical Circuits and Systems.* 14(4): 715–726.

[23] Zhao, S., J. Yang, and M. Sawan. 2021. Energy-Efficient Neural Network for Epileptic Seizure Prediction. *IEEE Transactions on Biomedical Engineering.* 69(1): 401–411.

[24] Li, F., B. Liu, X. Wang, B. Zhang, and J. Yan. 2016. Ternary Weight Networks. *arXiv Preprint.* arXiv:1605.04711.

[25] Birkl, C. 2017. *Oxford Battery Degradation Dataset 1*. Oxford: University of Oxford.

[26] Nagel, M., M. Fournarakis, Y. Bondarenko, and T. Blankevoort. 2022. Overcoming Oscillations in Quantization-Aware Training. In *Proceedings of the International Conference on Machine Learning (ICML)*. 16318–16330. PMLR.

[27] Gupta, K., and A. Asthana. 2024. Reducing the Side-Effects of Oscillations in Training of Quantized YOLO Networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2452–2461.

[28] Birkl, C. 2017. *Oxford Battery Degradation Dataset 1*. Oxford: University of Oxford.

[29] Kumar, P., G. Rankin, K. R. Pattipati, and B. Balasingam. 2022. Model-Based Approach to Long-Term Prediction of Battery Surface Temperature. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics.* 4(1): 389–399.

[30] Raijmakers, L. H. J., D. L. Danilov, R. A. Eichel, and P. H. L. Notten. 2019. A Review on Various Temperature-Indication Methods for Li-Ion Batteries. *Applied Energy.* 240: 918–945.

[31] Tang, X., K. Yao, B. Liu, W. Hu, and F. Gao. 2018. Long-Term Battery Voltage, Power, and Surface Temperature Prediction Using a Model-Based Extreme Learning Machine. *Energies.* 11(1): 86.