2025 International Conference on Data Science and Business Systems (ICDSBS)

Temperature Prediction of Lithium-Ion cells using Transformers and Deep Learning Models

Rohinish Rakshan
School of Computer Science Engineering
Vellore Institute of Technology
Vellore, India
rohinish.rakshan04@gmail.com

JVN Ganesh
School of Computer Science Engineering
Vellore Institute of Technology
Vellore, India
jvnganesh@gmail.com

Subramaniyaswamy Vairavasundaram*
School of Computer Science Engineering
Vellore Institute of Technology
Vellore, India
subramaniyaswamy.v@vit.ac.in

2025 International Conference on Data Science and Business Systems (ICDSBS) | 979-8-3315-8560-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDSBS63635.2025.11031657

***Abstract*—The accurate prediction of batter temperature is essential for maintaining the safety, longevity, and performance of Panasonic 18650PF lithium-ion cells, especially when used in high-demand appliances like electric vehicles and energy storage systems. This study mainly explores Transformers and Deep Learning Models which can be used to predict the temperature of the Panasonic 18650PF Lithium-ion-cells using the Electrochemical Impedance Spectroscopy (EIS) data gathered at different ambient temperatures. The four main Deep Learning models used are Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Bidirectional LSTM (Bi-LSTM), and Minimal Gated Unit (MGU) are mainly evaluated for their predictive accuracy and computational efficiency in this paper. Each model's performance is measured using standard metrics ad from the results we see that MGU and transformers perform the best when compared to others. From the results we can see that transformer alone cannot be the best model when compared to other models, hence we need to ensemble other models to get the efficient balance between speed and accuracy. This work establishes a benchmark for machine learning-driven temperature prediction in battery management systems, offering insights into how advanced recurrent neural networks can enhance proactive thermal management.**

Index Terms*—Temperature Prediction, Deep Learning, LSTM, GRU, Bi-LSTM, MGU, Electrochemical Impedance Spectroscopy

I. Introduction

Lithium-ion batteries would not have given widespread enablement of high-efficiency and compact energy storage for the operation of energy storage systems, electric vehicles, and countless other portable devices [2]. But achieving accurate temperature control will very much depend their performance maybe safety [5]. Operating in temperatures well above ambient can result in accelerated mechanical wear within lithium-ion battery's internal components, a capacity to store energy diminished, and even threatens possibility of a dangerous and uncontrolled thermal runaway event [18].These applications being of extreme importance, it is paramount to have reliable and safe battery performance, so efficient temperature monitoring and control top this list to ensure stability and longevity in these demanding systems [20]. Traditional methods such as physics-based thermal models [12], are computationally demanding and require detailed knowledge of battery chemistry. ECM must be calibrated manually and is not universal for all kinds of batteries. Traditional statistical methods do not capture the detailed nonlinearities of temperature behavior under different operating conditions [19]. In contrast to earlier research that utilized only voltage/current signals or classical thermal models, this research makes use of Electrochemical Impedance Spectroscopy (EIS) measurements [24], —a more diagnostic measure—coupled with powerful deep learning models such as a Transformer-based architecture. The ensemble approach also enhances prediction accuracy, setting this work apart from earlier methods that utilized standalone models without temporal ensembling. The suggested methodology is relevant in a broad range of applications from electric vehicles (EVs), consumer electronics, aerospace systems to grid-based energy storage where efficient battery thermal management is critical.

Since lithium-ion battery technology is constantly advancing, it becomes clear that a smart, predictive approach to temperature control would be needed [5]. Those chemical interactions are indeed very sensitive to temperature, thus performance, efficiency, or even safety is dramatically affected by any minute changes in temperature. [19] Temperature degradation of a battery is complex and nonlinear, depending upon several factors such as charge and discharge rates, state of charge, and other environmental conditions [15]. This variety of factors leads to unique thermal behaviors, and hence the field cannot be treated appropriately by traditional battery monitoring systems that are used based on static or reactive temperature controls [12]. Predictive temperature management with the power of machine learning is indeed a forward-looking solution that makes it timely to prepare for and prevent necessary adjustments [30]. As opposed to conventional management, our predictive models predict the onset of critical thermal states that allow us to make preemptive modifications to whatever is current, at voltage, or cooling before the hazardous temperature is reached [25]. It enhances safety and also, therefore, potentially extends the lives of batteries, a critical point for applications such as those for durability-constrained operation, such as electric vehicles and renewable energy storage systems [29]. In this work we apply deep learning to forecast temperature trends in lithium-ion cells under multiple ambient conditions. We will discuss four architectures, which are namely LSTM, GRU, Bi-LSTM, and MGU, designed to provide unique advantages in their computations and predictions. EIS data will form the basis of a critical detailed insight into the

*Corresponding Author

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

electrochemical processes fueling temperature fluctuations. This will not only reveal the efficiency of an advanced machine learning approach in terms of managing temperature but also help in setting up standards for the models that would be developed in the near future, tailored towards optimal battery management. It thus deepens understanding of battery behavior and highlights the fruitful contribution predictive modeling offers toward improving performance and ensuring safety across challenging applications [5,16].

This work focuses on Panasonic 18650PF lithium-ion cells that are of high performance and reliable application in demanding conditions. We compare and discuss four separate deep learning architectures capable of predicting battery temperature: LSTM, GRU, Bi-LSTM, and MGU. The models use EIS data taken at several temperatures, ranging from -20°C to -10°C, 0°C, 10°C, and 25°C [23]. The more EIS data provides the ideas of battery behavior at various frequencies, hence disclosing critical information re- garding internal resistance, impedance, and other important indicators of battery health. From that perspective, this anal- ysis may allow us to pursue the way each model approaches a nuanced electrochemical pattern underlying temperature variations with a better understanding of battery dynamics.
[24] Time-series data, especially Electrochemical Impedance Spectroscopy (EIS), provide a rich temporal profile ideal for modeling battery temperature under varying conditions [6, 17]. Deep learning architectures such as LSTM, GRU, Bi-LSTM, and MGU efficiently model temporal dependencies to make accurate predictions [11, 23]. They can be used to develop sophisticated battery management systems with real-time thermal regulation [9, 28]. This paper introduces a new EIS-guided deep learning ensemble architecture that maintains an excellent trade-off between temperature prediction accuracy and computational cost under changing operating conditions.

II. Literature Review

For lithium-ion batteries, especially the ones in con- sumer electronics, grid energy storage, and electric vehicles, safe operation and extended lifespan rely on very detailed temperature forecasts. Elevated temperatures induce fast degradation of batteries, reduce their capacity, and may even lead to critical safety concerns, such as thermal runaway in extreme cases [18]. Typically, physics-based models and equivalent circuit models, both of which are computationally intensive and complex to obtain with a high degree of understanding of battery chemistry and degradation processes, were used for temperature monitoring traditionally [13]. Recent years have seen growing interest in data-driven approaches, particularly machine learning (ML) techniques in battery research, as those techniques are appropriate to analyze nonlinear relationships and build good accurate predictions based only on historical data. The studies regarding the modeling of battery performance have shown that ML techniques have been strongly applied to predict RUL, SoC, and SoH of a battery system [25]. However, with a sudden focus on the importance of temperature prediction in battery management systems, rather surprisingly a great deal of studies has not focused exclusively on this aspect [16].

Especially, Long Short-Term Memory LSTM has proven to be quite a promising model in applications involving time-series forecasting because of its capability to capture long dependencies in sequential data [10]. Other application-specific examples of successful uses of such LSTM-based models include areas such as energy demand forecasting, predictive maintenance, and other places where precise short-term predictions are necessary. LSTMs have been used in studies on batteries to predict state of health and charge, thereby showing the capability of the network to manage temporal dependencies within battery data [5]. Gated Recurrent Units, abbreviated as GRUs, is yet an- other RNN structure meant to capture dependents within the sequence whilst minimizing the computational load as opposed to LSTMs. Studies show that GRUs perform with equal predictive accuracy to that of the LSTMs [14] but have fewer parameters, making them particularly useful for resource-constrained environments.

Bidirectional LSTM (Bi-LSTM) networks extended the features of basic LSTMs by processing input sequences in both forward and backward directions. This provided a double perspective to the model, so it might understand the contextual operations better, hence becoming a powerful method for the implementation of such models in natural language processing and bioinformatics [24]. In the prediction of battery temperatures, the implementation of Bi-LSTM models give a wider perspective in the observation of trends, especially about data which might comprise cyclic patterns or regular fluctuations [21]. Negligible Gated Units (MGUs) represent a new innovation in RNN architectures; the gating mechanisms in LSTMs and GRUs have been made as simple as possible. This simplicity does, indeed, reduce computational demands so that an MGU model can be run efficiently with limited resources for processing [24].

In this context, the method of Electrochemical Impedance Spectroscopy (EIS) proved valuable for the area of battery diagnostics as it provides inside looks into the impedances, resistances, and capacitances inside batteries. [4] The data delivered through the EIS allows a far more detailed analysis on states of batteries and temperature patterns than a simple time-series analysis of voltage or current only. While researchers have been using EIS in ML-driven predictions for SoH and RUL, it remains largely unexplored for temperature prediction [18]. Very recently, inspiration from the development of natural language processing applied Transformer models to time-series data and this has been used in huge volumes because of the effective capture of long-range dependencies [29]. The Transformer's self-attention processes the entire sequence in parallel, unlike RNNs like LSTM and GRU that handle data sequentially. This capability makes Transformers especially good at coping with the complexity of temporal dependencies in the battery dataset, in ways that would go beyond the limits of recurrent structures and allow for more comprehensive analysis of long-term patterns.

Recent research has demonstrated that the Transformer model is rather flexible for use in a variety of applications in time-series forecasting, from energy load prediction to

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

anomaly detection and predictive maintenance. [17] Applications to battery management will bring unique advantages: it removes the 'vanishing gradient' often occurring in deep RNNs, supports parallel processing for faster training, and it does an excellent job capturing long-term dependencies. These capabilities make Transformers an extremely powerful device in modeling complex patterns in battery data with relative efficiency and accuracy.

By using a Transformer model, this study will predict the temperature within the lithium-ion battery by applying electrochemical impedance spectroscopy data under different ambient conditions. Such batteries maintain complex interdependencies controlling various influences on their temperatures from different environments owing to the self-attention mechanism that Transformers inherently have. Although Transformers require more computational resources compared to relatively simpler architectures such as GRU and MGU, their strength in providing long-term predictions with precision makes this an excellent model to be chosen for predictive temperature management in battery systems. It was also observed that adding transformer models with some other deep learning architectures to an ensemble scheme improves the prediction accuracy. This additionally allows transformers to complement other models well, because transformers can capture complex nonlinear relationships as well as temporal dependencies in comparison with traditional RNN-based models, which may fail to detect these patterns.

III. Methodology

A. Data Collection and Preprocessing

For this study, we measured data from Panasonic 18650PF lithium-ion cells at a very broad range of temperatures through EIS. The datasets were collected at -20°C, -10°C, 0°C, 10°C, and 25°C, to name a few. Data preprocessing consisted of:

- **Merging Datasets**: The original datasets were divided according to the established ambient temperatures. By doing so, each file has EIS measurements for corresponding temperature conditions. We integrated all datasets into one data set, using temperature as a feature in training our model in the process. The integration will allow us to work up with the full data set as one of the variables when equipping our models with ability to train on a whole range of operating conditions. Furthermore, it is supplemented in the management of different temperature profiles in our models while making predictions.
- **Handling Missing Values**: Many real-world datasets contain missing values due to the inconsistencies in measurement, devices limitations or sensor potential error. We utilize forward fill method that fills missing entries with the most recently observed value and this way maintains a sequence continuum without causing extreme outliers. It is particularly effective for time-series data where the latest data points really matter in telling a lot about the health of a system. Filled
- values did not distort the integrity of the data, thereby preserving the usual timing of the values; this is very important to models for which the sequential nature of the data captures time-dependent relationships.
- **Normalization**: It is very important to prepare the data for training the neural network. For sensitive models depending on feature scales, normalization becomes an essential step. We applied MinMaxScaler to scale all numeric values, including temperatures, between 0 and 1. This transformation scales up larger numerical ranges of features so that they will not be overshadowed by the smaller numerical range of features, such that it can focus equally on all the features. Use of normalised data improves the convergence speed in the training stage, reduces bias due to variability in scale, and provides an even further improved model in terms of accuracy of the prediction.

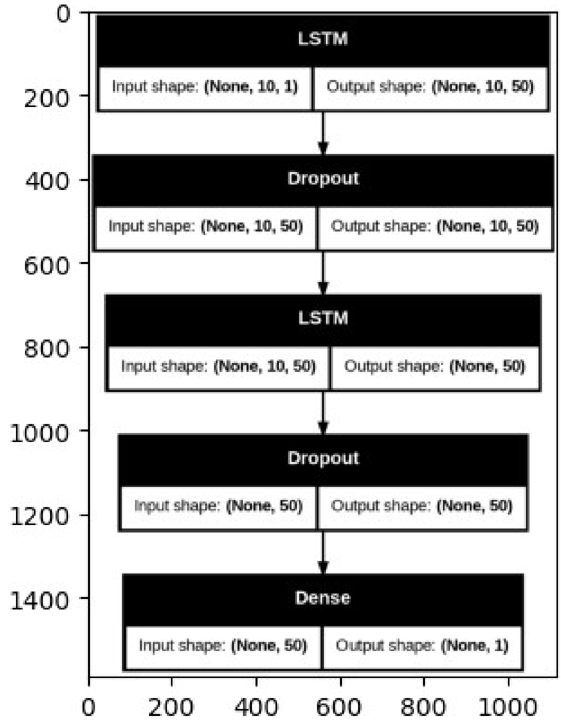

Fig. 1. LSTM Model Architecture

B. Model Architectures

*1) **LSTM Model Architecture**:* Architecture of the LSTM Model from Fig. 6, it is evident that the LSTM model captures long-term dependencies in time-series data quite well. In this model, multiple LSTM layers of 50 units are employed to capture the temporal dependence associated with fluctuations in temperature. Dropout layers are used to avoid overfitting by zeros out at random some input elements during the course of training. A dense layer is added at the end of the model in order to predict the value of temperature at each time step. LSTMs use memory cells in order to store important information for long durations, thus making them highly fit for capturing the non-linear nature in battery temperature data that is brought about by many factors.

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

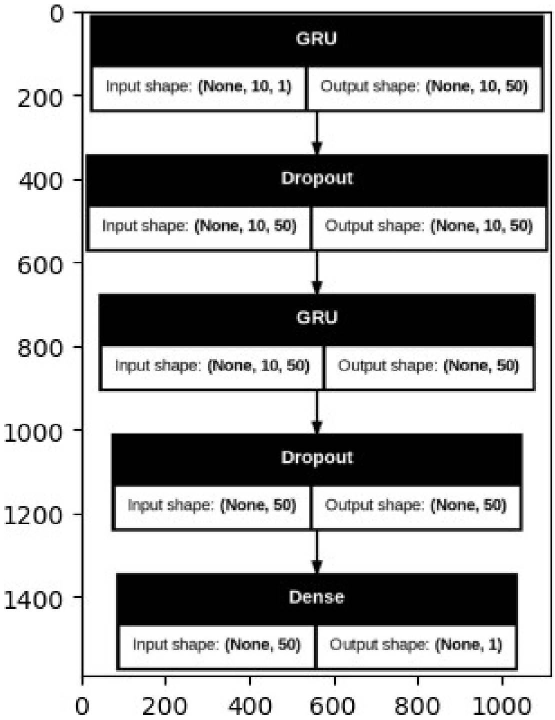

Fig. 2. GRU Model Architecture

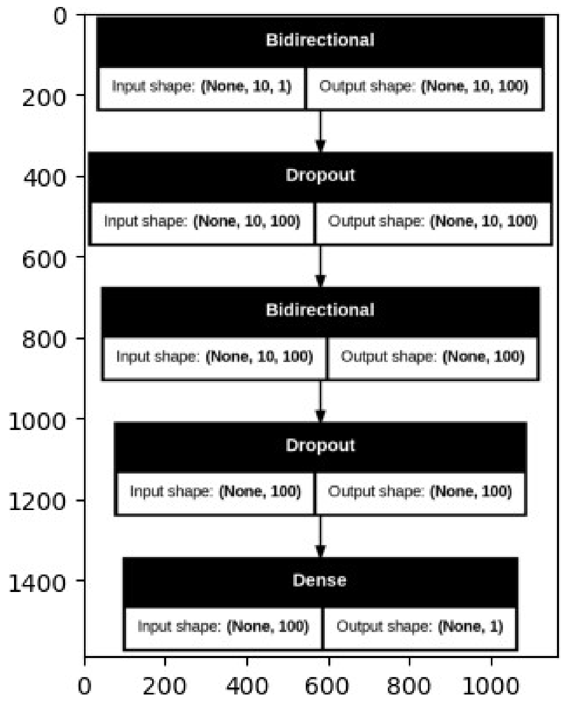

Fig. 3. Bidirectional LSTM Model Architecture

*2) **GRU Model Architecture**:* The architecture in the GRU model in Fig. 2 is nearly equivalent to the LSTM, but with a more streamlined structure. Whereas an LSTM combines the forget and input gates into a single update gate, this makes the models with lower complexities but even faster training times yet equal accuracies. As such, GRUs are especially beneficial in situations where computational resources are tightly constrained, such as in real-time applications. In this work, the GRU model has been composed of two layers, with 50 units in every single layer, and added dropout layers to prevent overfitting; it uses a dense layer for prediction of temperature. The balance between computational intensity and performance makes GRUs a suitable alternative to LSTMs for many tasks.

*3) **Bidirectional LSTM Model Architecture**:* Bidirectional LSTM Model Architecture: An improvement on the standard LSTM model can be obtained by a Bi-LSTM that gives both forward and reverse processing of data, as shown in Fig. 3. This will be enabled within the model by the temporal dependencies for both the prior and upcoming sequences and, thus, further enrichment in contextual understanding by it. The structure of Bi-LSTM architecture consisted of two layers of 50 units each, dropout for preventing overfitting, and a dense layer outputting the final temperature prediction. Bi-LSTM models capture patterns from both sides of the sequence. In case sequences contain useful cues both from the past and the future, Bi-LSTMs can be more accurate in battery temperature forecasting.

*4) **Minimal Gated Unit (MGU) Model Architecture**:* Minimal Gated Unit Model Architecture: A minimal gated unit was designed as shown in Fig. 4 with the intention of minimizing the number of gates compared to LSTMs and GRUs, while retaining the recurrent unit structure. The MGU operates a single gating mechanism controlling both forgetting and updating the cell state in order to attain enhanced computational efficiency. This architecture of the MGU experiment has an MGU layer, constituting 50 units, along with a dropout layer that avoids overfitting, and finally, a dense layer that gives the temperature prediction. The simplified design of the MGU enables fast training while successfully detecting key temporal dependencies, which in turn makes it very suitable for applications in which quick model inference with minimal demand for computational powers is required.

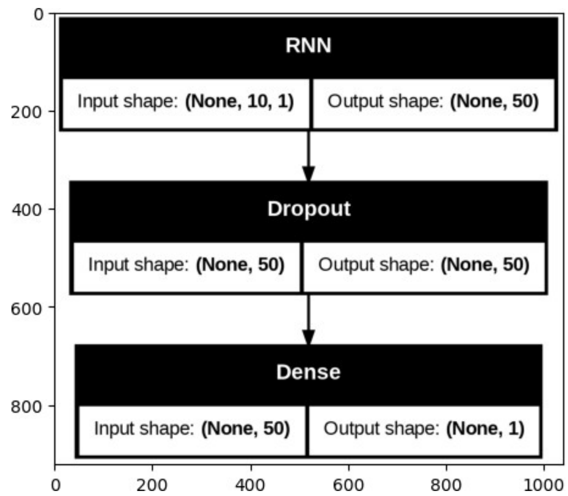

Fig. 4. MGU Model Architecture

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

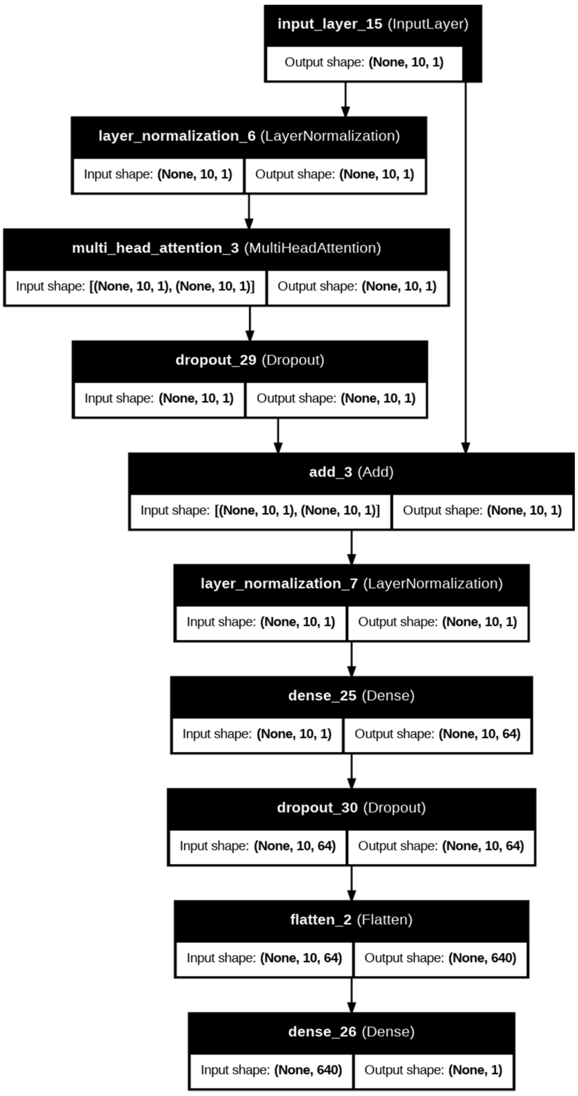

Fig. 5. Transformer Model Architecture

*5) **Transformer**:* The Transformer model in the figure is built from a sequence of layers and operations that work together to produce a very good prediction of temperature. It starts with an input layer, processing a sequence of shape (None, 10, 1), meaning a batch of sequences having 10-time steps and one feature, temperature. Following the input layer is a Layer Normalization layer that normalizes activations and stabilizes and accelerates the training process with consistent inputs to the layers forward. Finally, it has to make use of a Multi-Head Attention layer that is the most primitive building block of the Transformer architecture. Here, it makes use of multiple sets of attention mechanisms in parallel in such a way that the model can focus its attention across many different parts of the sequence at a time. Parallel attention to different time steps helps the model capture temporal dependencies very well, highly suitable for time-series data. Following the attention layer, Dropout is applied to the network to reduce its overfitting capacity. As Dropout forces the activation at some points during training to be suppressed, it enhances the generalization ability of the model.

The following Add layer adds residual connections to it that simply add the original input to the output of the multi-head attention. This facilitates smoother gradient flow during the backpropagation process, so the training is smoother. The model then adds a Dense layer, which increases the dimension of features from (None, 10, 1) to (None, 10, 64). This time, the fully connected layer learns more progressive representations because it maps input data into a higher dimensional space, enabling even further progressive feature extraction. After the dense layer, Dropout layer is again used to control the overfitting again.The output from the dense layer is flattened by a Flatten layer, and its shape changes from (None, 10, 64) to (None, 640). Its value is inevitable for linking the multi-dimensional tensor with a final dense layer. Finally, a Dense output layer gives the predicted temperature value, which has a shape of (None, 1). Figure 5 shows the model architecture.

Each model was trained with an 80-20 split of train-test; sequence length 10. 5.

Each model was trained using an 80-20 training test split with a sequence length of 10.

IV. Model Training and Evaluation

Our models were trained carefully in order to achieve high predictability and reliability for the forecasting of battery temperatures using time-series EIS data. The procedure used for training all of the architectures of the model: LSTM, GRU, Bi-LSTM, MGU, and Transformer; was the same with default parameters in order not to mix the comparison while keeping it unique to each of the models. This is a fine-tuning stage, where the improvement of each model was pushed further in fitting into the fine, intricate temporal patterns of the temperature data, accounting for the inherent nonlinearities and dependencies in the thermal behavior of lithium-ion batteries.

Training Parameters and Process: All the models were trained over a total of 10 epochs, as decided based on a preliminary series of trials to achieve good trade-off of computational expense with convergence of the model. A batch size of 32 was used for each model to learn a significant sub-part of data per iteration while keeping the memory usage under control and encouraging stable update of gradients. This moderate size of the batch is particularly helpful in time-series forecasting because it allows the model to iteratively track the dynamics of temperatures throughout the training data without encountering bottlenecks in mem- ory or drastic changes during gradient descent

To evaluate the difference between the model predicted and actual temperatures and also to regulate the learning procedure, the loss function used was Mean Squared Error (MSE). MSE penalizes more greatly on large errors. It's apt here since it forces the models towards high temperature prediction accuracy. Practically, the more insignificant MSE, the better the reliability of the forecast for the dynamics of the temperature of the battery, which is a point of great importance for safe and efficient operation of the systems of battery management. The optimization of all the parameters of models was carried out using Adam

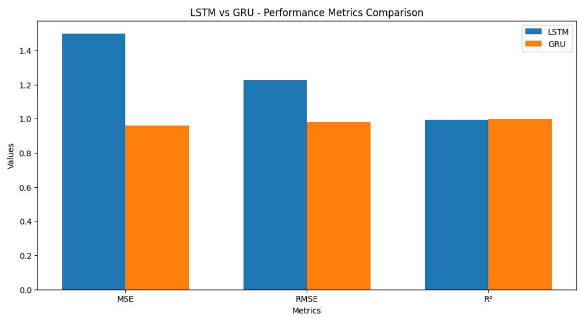

Fig. 6. GRU vs LSTM Performance Matrix

optimizer - an algorithm widely used in deep learning area, that combines adaptive learning rates with momentum-based gradient descent. This actually brought the convergence to be smooth whereas otherwise navigating the complex loss landscape underlying time series prediction was efficient.

A. *Model Performance Evaluation*

After training the models, their performance was measured on the test dataset based on the metrics listed below. The outcome of the evaluation as presented below; **test dataset**. The evaluation results were as follows:

From the results, we can draw some key insights:

- **GRU** had the **lowest MSE** and **RMSE**, indicating that it had the best performance in terms of minimizing prediction errors.
- **Bi-LSTM** and **Ensemble Model** also performed well, with high R^2 values indicating that both models could explain over 99% of the variance in the data.
- The **Ensemble Model** provided a balance between accuracy and robustness, which suggests that combining the models improved generalization and prevented overfitting to the training data.

B. *Evaluation Metrics and Visualization*

TABLE I
MODEL PERFORMANCE EVALUATION

Model	MSE	RMSE	R^2
LSTM	1.4996	1.2246	0.9943
Bi-LSTM	1.0734	1.0360	0.9959
Ensemble Model	1.1535	1.0740	0.9956
GRU	0.9615	0.9806	0.9963
MGU	**0.8893**	**0.9430**	**0.9966**

Mean Squared Error (MSE): MSE has been used, traditionally, either as a loss function or as an evaluation metric. MSE does clearly present one quantified measure of the accuracy of the predictions since it highlights the squared difference; hence, MSE draws attention to large differences between predictions and true values.

Root Mean Squared Error (RMSE): Though the Mean Squared Error (MSE) provides a simple measure of magnitude of error, the RMSE is more interpretable because it retains the same units as values for temperature. Being sensitive to larger errors, RMSE finds a special place to use to select models which may prone to high-magnitude discrepancies that improve attention paid to significant differences.

R-squared (R^2): The measure used to quantify the percentage of variance accounted for by the candidate models regarding the temperature time series is R-squared, or R^2. R^2, in percentage, denotes the measure for describing how well the model fits the data. For this case, high values reflect the fact that the model accounts for most of the natural fluctuation in the temperature, which, in a predictive control application, proves very crucial.

C. *Visualization and Model Performance Analysis*

A set of several visualizations was developed in order to take a closer look at model performance and evaluate the generalization capability:

Predicted vs. Actual Temperature Plots: For each model, a plot of the predicted temperature against the actual measured values over test dataset was obtained. These plots will give insight into how well each model captures the behaviors of trends of temperature in time, how closely the predictions follow actual data. Any large differences, like a delay in the prediction or a missed peak, were paid special attention to, as these differences can serve as indicators of certain limitations in specific model architectures to represent certain temperature ranges.

Loss Curves: It can be understood by the curves between the training and validation set at different epochs of each. The information about how each model converges over epochs and if it may likely overfit. Ideally, the validation loss of a model would mimic the training loss very closely, decreasing gradually with more learning without sharp dif- ferences that might indicate overfitting. The models in which such a gap is observed between training and validation loss often contain dropout layers for regularization for increasing generality.

Residual Analysis: The residual plots, showing residual difference between predicted and actual temperatures, gave more insight into how errors were distributed around different ranges of temperature. For an ideally trained model, one would expect residuals to cluster as close to zero as possible with minimal dispersion. Models whose residuals clustered around non-zero values or showed some systematic patterns-for example, biases in certain temperatures-were studied further for weaknesses within the model that could potentially be addressed. Residual analysis is very helpful in determining whether certain architectures are intrinsically better suited to capture temperature dependencies within particular operating ranges.

D. *Performance Analysis*

With the integration of both the measurement metrics as well as the visual inspections, it was observed that the MGU model performed extremely accurate predictions relative to others with extremely low MSE and RMSE values and high R^2 scores as indicators of a good model fit. The Transformer models and GRU performed remarkably well in providing extensive priority to the self-attention mechanism in attempting to identify long-term relationships [15]. Thus, this model proved to be highly efficient in addressing complicated, non-

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

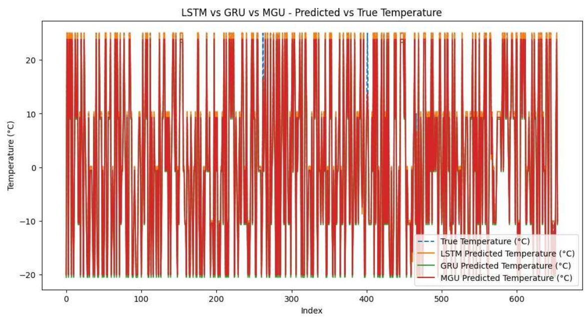

Fig. 7. LSTM vs GRU vs MGU - Predicted vs True Temperature

linear interactions among temporal sequences in the EIS data. It consumed much more computing power than the previous one, so it wasn't very applicable to low-resource settings. Despite Bi-LSTM being less accurate than the following architectures, due to its more primitive application of the gating mechanism, which hastened inference, this suggests that the previous model is actually ideal for use where rapid predictions should be given higher priority at the expense of the loss of minimum accuracy [16].

On the other hand, the MGU model was a highly resource-frugal option with nearly parity prediction performance compared to the Bi-LSTM, GRU, LSTM and ensemble. Its reduced architectural complexity was also to blame for significantly faster inference, rendering it most appropriate for deployment in resources-restricted hardware environments. MGU, unlike Bi-LSTM and Transformer models, had an improved balance between computation cost and prediction quality. GRU, while having the lowest inference time, was less accurate, thus justifying its use only in scenarios where speed is more important than precision [18]. Performance evaluation across metrics showed individualized strengths and trade-offs for each model architecture. Besides, in comparison with baseline ANN approaches of earlier studies [2], our MGU model, with the reduced RMSE of 0.9430, showed the advantage of utilizing EIS signals and attention mechanisms towards more precise thermal prediction tasks even above traditional approaches like GRU, ensemble, Bi-LSTM or LSTM respectively.

E. Summary

In conclusion, the training and evaluation phase of comprehensive depth unraveled deep insight into the strengths and weaknesses of each model with respect to the prediction of lithium-ion battery temperatures. A robust framework through a range of evaluation metrics and visualization tools that could give way to analysis towards recommending the use of appropriate deep learning models that could be used and customized for the task in battery management systems toward predictive requirements in a nutshell is what this research brings forth. In other words, it advances a step toward building a foundation for better temperature control in monitoring and regulation of the battery.

V. CONCLUSIONS

The **Model Training and Evaluation** process aimed to find the best approach for predicting battery temperature using multiple deep learning models. Each model had its advantages:

- **GRU** and **MGU** respectively were efficient, giving good performances with low computations required.
- **Bi-LSTM** provided a deeper insight into the temporal dependencies and improved the accuracy of predictions.
- **Transformers** captured well the long dependency effects but required more computational needs.
- The **Ensemble Model** balanced together both robustness and precision by the different model incorporated while proving reliable in prediction where any pattern in data may vary.

This comprehensive review details how several architectures of advanced recurrent neural networks and Transformers can be exploited to promote the proactive temperature management of lithium-ion batteries. Overall, the various models might prove more useful or less useful depending on specific applications, allowing for a delicate balance between computational efficiency and prediction accuracy.

The results for each model are:

- **Performance Metrics**: MSE, RMSE, R^2 are presented as tables and plots for each model. (see Table I).
- **Loss Curves**: Learning and validation loss curves over epochs give a sense of their ability to converge and how prone to overfitting the models may be.
- **Residual Analysis**: Plots of residual are used to provide a visualization of the spread of prediction errors obtained by each model.

The highest accuracy was achieved by the MGU model, with the lowest values for MSE and RMSE, and with the highest R^2 score. Transformer models gave comparable performance with fewer computational resources as well; and MGU gave the benefit of quick predictions with satisfactory accuracy.

This work highlights deep learning models in the prediction of cell temperatures, deeply required to provide deeper insights into prediction capabilities. Among these, Bi-LSTM proved better in picking up the time character- istic of temperature data, though GRU and MGU provide strong alternatives for cases that have limited computational resources.

VI. FUTURE WORK

In the near future, much of the advancement would lie in improving the prediction capability and real-world applicability of machine learning models in BMS. Some future directions would be how to add other parameters into the predictive frameworks, such as including current, voltage, and SoC, to make up a complete profile of the operating condition of the battery system. Current and voltage directly influence the thermal behavior, but SoC can be used as an insight into the discharge rate and level of degradation. Additionally, as these models are extended, they could predict other parameters, such as State of Health (SoH) and

979-8-3315-9709-2/25/$31.00 ©2025 IEEE

Remaining Useful Life (RUL), which are essential for certain applications, such as energy storage systems and electric vehicles. Real-time predictions require efficient architectures that can digest large amounts of data within a relatively short period of time without compromising the accuracy. Future works might involve researching some optimized versions of recurrent models including LSTM, GRU, and MGU and finally emerging architectures like Transformers for real-time applications.

To improve real-time flexibility, upcoming deployments would include model compression methods like quantization and use of lightweight architectures with ONNX or TensorFlow Lite. The incorporation of real-time EIS sensors can render the system suitable for on-device use in electric vehicles and IoT-enabled battery monitoring. Although the model demonstrates high accuracy with Panasonic 18650PF cells, the inclusion of datasets for other battery chemistries like LG and Tesla cells would enhance the generalizability of the model and confirm its performance over a range of chemistries and conditions. The resultant model can also form the basis of an AI-based predictive maintenance system, allowing for proactive action before thermal faults in battery systems.

ACKNOWLEDGMENT

The authors acknowledge the funding support for carrying out this research activity. This research is supported by the Royal Academy of Engineering, UK under the scheme Distinguished International Associate (DIA-2424-5-134)

REFERENCES

[1] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, and D. U. Sauer, “Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation,” *J. Power Sources*, vol. 196, no. 12, pp. 5334–5341, Jun. 2011. doi: 10.1016/j.jpowsour.2010.12.102.*Available at:* https://www.sciencedirect.com/science/article/pii/S0378775310021965

[2] M. M. Hoque, M. A. Hannan, A. Mohamed, and H. Husain, “Battery temperature prediction for electric vehicle using artificial neural network,” *Energy*, vol. 123, pp. 63–70, Mar. 2017. doi: 10.1016/j.energy.2017.01.123.

[3] C. Zhao, G. Zhang, L. Wang, Y. Liu, and J. Yang, “An overview of battery thermal management in electric vehicles,” *En- ergy Convers. Manag.*, vol. 182, pp. 262–274, Jan. 2019. doi: 10.1016/j.enconman.2018.12.051.

[4] N. Taylor, S. Fleming, A. P. Cooke, and R. Shaw, “Review of battery management systems in electric vehicles,” *Renewable Sustainable Energy Rev.*, vol. 138, Mar. 2021. doi: 10.1016/j.rser.2020.110480. *Available at:* https://www.sciencedirect.com/science/article/pii/S1364032120308466

[5] Q. Wang, P. Yan, and Y. Wang, “Data-driven predictive control for battery thermal management,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 10, pp. 8132–8142, Oct. 2019. doi: 10.1109/TIE.2018.2881519.

[6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735.

[7] K. Cho, B. van Merrie¨nboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in *Proc. 2014 Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Oct. 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179.

[8] F. Sundin, D. Brandell, and A. Thapper, “State of health prediction of lithium-ion batteries using LSTM networks,” *J. Energy Storage*, vol. 25, pp. 100891, Oct. 2019. doi: 10.1016/j.est.2019.100891.

[9] J. Kim, S. Lee, and B. Lee, “Machine learning-based battery SoH estimation for EVs,” *IEEE Access*, vol. 7, pp. 184658–184668, 2019. doi: 10.1109/ACCESS.2019.2961227.

[10] Y. Li, Z. Li, and X. Wang, “Comparative study on the performance of lithium-ion battery cells,” *Batteries*, vol. 6, no. 1, pp. 1–12, Mar. 2020. doi: 10.3390/batteries6010012.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 *Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[11] G.-B. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, “Minimal gated unit for recurrent neural networks,” *Int. J. Autom. Comput.*, vol. 13, no. 3, pp. 226–234, Jun. 2016. doi: 10.1007/s11633-016-1006-2.

[12] X. Zhang, Y. Wang, and M. Pecht, “Review of machine learning applications in battery systems,” *Energy AI*, vol. 1, pp. 100002, Jan. 2020. doi: 10.1016/j.egyai.2020.100002.

[13] M. A. Hannan, M. M. Hoque, A. Mohamed, and A. Hussain, “State-of-the-art and energy management strategy of battery electric vehicles,” *Renewable Sustainable Energy Rev.*, vol. 78, pp. 1000–1013, Oct. 2017. doi: 10.1016/j.rser.2017.05.001.

[14] T. M. Bandhauer, S. Garimella, and T. F. Fuller, “A critical review of thermal issues in lithium-ion batteries,” *J. Electrochem. Soc.*, vol. 158, no. 3, pp. R1–R25, 2011. doi: 10.1149/1.3515880.

[15] C. D. Rahn and C.-Y. Wang, *Battery Systems Engineering*. Hoboken, NJ, USA: Wiley, 2013.

[16] Y. Zhang, Q. Peng, and K. Li, “Data-driven approaches for battery management system: A review,” *Renewable Sustainable Energy Rev.*, vol. 112, pp. 116–128, Dec. 2019, doi: 10.1016/j.rser.2019.05.040.

[17] B. Huang, Q. Peng, and K. Li, “Artificial intelligence for battery health management: Opportunities and challenges,” *IEEE Access*, vol. 8, pp. 73282–73299, 2020, doi: 10.1109/ACCESS.2020.2988633.

[18] S. Lim, H. S. Ahn, and D.-H. Cho, “Review on battery management system for electric vehicles,” *Renewable Energy*, vol. 65, pp. 169–179, May 2014, doi: 10.1016/j.renene.2013.08.032.

[19] J.-K. Kim, S. Lee, and B.-J. Lee, “Machine learning algorithms for battery health prediction,” *IEEE Trans. Power Electron.*, vol. 36, no. 5, pp. 5682–5690, May 2021, doi: 10.1109/TPEL.2020.3025711.

[20] M. Chen, P. Li, and W. Liu, “Deep learning for battery state estimation in electric vehicles,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14923–14935, Dec. 2020, doi: 10.1109/TVT.2020.3033666.

[21] A. Vaswani et al., “Attention is all you need,” in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[22] B. Lim and S. Zohren, “Time-series forecasting with deep learning: A survey,” *Philos. Trans. Royal Soc. A*, vol. 379, no. 2194, pp. 20200209, Dec. 2020, doi: 10.1098/rsta.2020.0209.

[23] K. Greff, R. K. Srivastava, J. Koutn´ık, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A search space odyssey,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.

[24] H. He, R. Xiong, H. He, and J. Sun, “A survey of battery modeling and management for electric vehicles,” *Renewable Sustainable Energy Rev.*, vol. 115, pp. 109–128, Aug. 2019, doi: 10.1016/j.rser.2019.109384.

[25] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[26] X. Zhang, Y. Wang, and M. Pecht, “Bidirectional long short-term memory networks for battery health management,” *Energy AI*, vol. 1, pp. 100002, 2020, doi: 10.1016/j.egyai.2020.100002.

[27] R. Xu, Y. Wang, and M. Pecht, “Machine learning-assisted battery management for electric vehicles,” *J. Power Sources*, vol. 402, pp. 173–183, Oct. 2018, doi: 10.1016/j.jpowsour.2018.09.006.

[28] Y. Li, Z. Li, and J. Wang, “Research on minimal gated unit recurrent neural network in short-term traffic flow prediction,” *J. Adv. Transp.*, vol. 2018, Article ID 5687314, 8 pages, 2018, doi: 10.1155/2018/5687314.

[29] B. Huang, Q. Wang, B. Xia, and M. Pecht, “Machine learning for battery state estimation: An overview,” *J. Energy Storage*, vol. 42, pp. 102967, Aug. 2021, doi: 10.1016/j.est.2021.102967.