

State of Health Estimation of Lithium-Ion Batteries Using Support Vector Regression and Long Short-Term Memory

Inioluwa Obisakin¹, Chikodinaka Vanessa Ekeanyanwu²

¹Ingram School of Engineering, Texas State University, San Marcos, USA

²Department of Geography and Environmental Studies, Texas State University, San Marcos, USA

Email: ifo2@txstate.edu, cve9@txstate.edu

How to cite this paper: Obisakin, I. and Ekeanyanwu, C.V. (2022) State of Health Estimation of Lithium-Ion Batteries Using Support Vector Regression and Long Short-Term Memory. *Open Journal of Applied Sciences*, 12, 1366-1382.

<https://doi.org/10.4236/ojapps.2022.128094>

Received: May 31, 2022

Accepted: August 16, 2022

Published: August 19, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Lithium-ion batteries are the most widely accepted type of battery in the electric vehicle industry because of some of their positive inherent characteristics. However, the safety problems associated with inaccurate estimation and prediction of the state of health of these batteries have attracted wide attention due to the adverse negative effect on vehicle safety. In this paper, both machine and deep learning models were used to estimate the state of health of lithium-ion batteries. The paper introduces the definition of battery health status and its importance in the electric vehicle industry. Based on the data preprocessing and visualization analysis, three features related to actual battery capacity degradation are extracted from the data. Two learning models, SVR and LSTM were employed for the state of health estimation and their respective results are compared in this paper. The mean square error and coefficient of determination were the two metrics for the performance evaluation of the models. The experimental results indicate that both models have high estimation results. However, the metrics indicated that the SVR was the overall best model.

Keywords

Support Vector Regression (SVR), Long Short-Term Memory (LSTM) Network, State of Health (SOH) Estimation

1. Introduction

The transportation industry is one of the leading causes of air pollution and ozone layer depletion. As a relatively convenient means of transportation in daily life, modern vehicles carry increasingly high requirements and expectations

from society on energy efficiency, reduction in emission, and environmental protection. Due to these reasons, more and more countries are investing a lot of resources both human and financial into the research and development of Electric Vehicle (EV) technology to achieve clean energy goals and resolutions. Electric vehicles, as a new transportation tool, have been highly valued by various countries and regions due to their advantages of energy-saving, zero emissions, and low pollution, with increasing efforts in their technological research and development [1].

EVs employ the use of a clean energy storage system to power a vehicle. This usually means large numbers of battery cells ranging from hundreds to thousands are connected in series/parallel to enable powering the vehicle. Due to positive attributes such as high specific energy, long cycle life, very low self-discharge rate, ideal-temperature range, and resultant minimal pollution to the environment, lithium-ion power batteries are the most widely used type of batteries in this sector. However, continuous usage of an Electric Vehicle leads to an increase in the lithium-ion batteries' charging and discharge cycles, resulting in the electrode materials gradually becoming inactive and leading to the performance degradation of the battery. Therefore, it is necessary to manage the batteries to ensure proper operations by a battery management system (BMS). Prediction of a battery's current state of health is one of the most critical issues of the battery management system [2].

According to statistics, the inaccurate State-of-health (SOH) estimation and life prediction of lithium-ion batteries are the leading causes of electric vehicle spontaneous combustion accidents [1]. Therefore, accurately estimating the lithium-ion battery state of health has become a research focus for many scholars. This has resulted in several research works where learning models such as Artificial Neural Networks, Recurrent Neural Network have been applied to predicting the health status of batteries. These works have focused on several battery components such as Internal Resistance measurements, Open Voltage capacity, and Voltage drop rate and how these components affect the SOH. However, this report focuses on accurately using machine and deep learning models to predict the state of health of batteries by using features extracted from discharged battery data.

2. Literature Review

The SOH reflects the general condition of a battery and its ability to deliver the specified performance compared to an unused or fresh battery. It is defined as the ratio of the full charge capacity of a battery in the current state and the full charge capacity of a battery when it is initially bought (Nominal Capacity).

$$\text{SOH} \approx \frac{\text{Current actual capacity (Ah)}}{\text{Nominal Capacity (Ah)}} * 100\% \quad (1)$$

Usually, the end of life is determined when the actual battery capacity is lower than the Acceptable Performance Threshold (APT) [2] [3]. APT is usually 70%

or 80% of rated capacity.

SOH is a subjective measure in which researchers have derived different definitions using varieties of different measurable battery performance parameters such as current, voltage, resistance, temperature, self-discharge rate, stress, strain, etc. Though SOH is a function of such parameters (that is they all affect the capability of the battery), it is generally expressed in terms of capacity, considering other parameters constant or keeping them unchanged during the moment [4]. An accurate estimation of SOH is important to forecast batteries' reliability, efficiency, and power delivering capacity and proper operation of the system [5].

It has been reported that capacity, internal resistance, power fade, and cycle life change with battery's age and hence these parameters are useful in predicting the behavior of the cell or battery [6]. Aging processes of a battery are irreversible changes in the characteristics of the electrolyte, anode, and cathode and the alteration in the structure of the components used in the battery. Battery aging can be divided into cycle aging and calendar one [7]. Cycle aging associates with the impact of battery utilization periods, and the calendar aging associates with the consequences of battery storage. Aging is considered for the estimation of SOH as it is highly related to change in capacity, internal resistance, and power fade [8]. Changes of these parameters help the researchers to find out which could be the best parameter for SOH estimation in accordance with the situations. For example, changes in the performance of battery's external behavior due to loss of rated capacity or due to an increase in temperature because of internal changes like corrosion.

Formulation of battery modeling is necessary to relate the battery parameters such as charging and discharging voltage, cycle life, temperature, etc. Battery modeling is divided into electrochemical models, empirical models and equivalent circuit models (ECMs). In the empirical model, the formulation of a model is based on the experimental data obtained from the batteries where we do not completely know the internal information of the battery activity. In order to predict the unknown information of the battery, some methods such as Kalman filtering (KF), fuzzy logic, neural networks (NNs), etc. are used to build the empirical model. In the electrochemical model, the models are based on the chemical processes that take place inside the battery. The electrochemical models are more accurate; however, these models are complex to analyze [9]. In order to reduce the complexity of the model, some reduction models such as single particle model are used [9]. In the fusion model, the models are based on combination of empirical and electrochemical models. Data are obtained from the finite element simulation (electrochemical models) of the battery phenomenon. The quantification of data is then obtained by building empirical models using methods like fuzzy logic, Kalman filtering, neural networks, etc.

Some researchers explain the battery health monitoring models/methods in different ways [10]. For example, Berecibar *et al.* [11] divided the methods of

SOH estimation into two parts: experimental technique and adaptive method. In experimental technique, previous data were considered, whereas in adaptive technique, some parameters were introduced which had been sensitive to degradation or aging of the battery.

Generally, models work accurately when used offline. However, the models do not work well when used in real-time and online; in this way, it is challenging to model the SOH or SOC for the entire battery pack compared to a single battery or cell. Therefore, designing the best model considering all the necessary parameters is essential. In this project, two approaches to predicting the SOH of Lithium-ion batteries are considered. In the first approach, the deep learning LSTM model is used in a time series format while the second approach focuses on exploring the regression ability of the machine learning SVR model. This is because the paper focuses on performances of both the machine and deep learning model with the eventual goal of observing the overall best model.

3. Methodology

The development of the overall learning model for SOH estimation consists of 2 major sections. The block diagram depicting the two major sections of the prediction process is shown in **Figure 1**. The first section contains the Data acquisition,

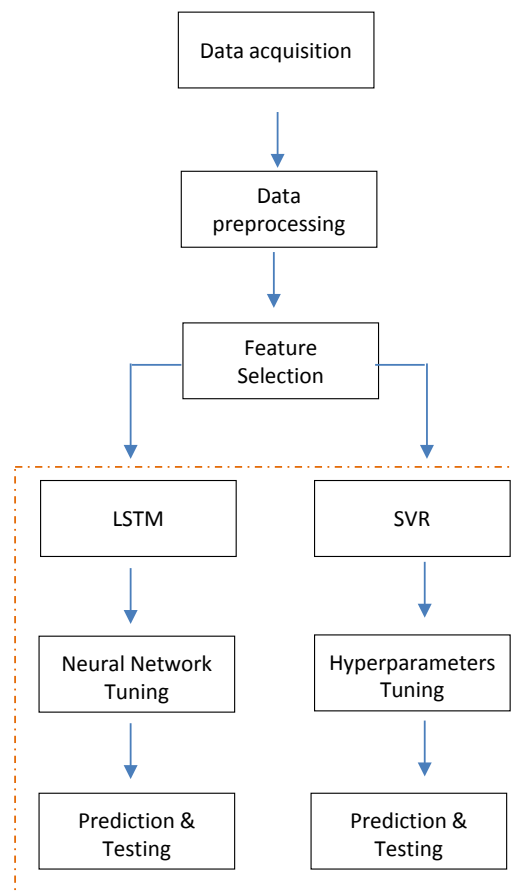


Figure 1. Learning model process flow.

preprocessing of data, feature Extraction. The second section is divided into two separate halves depicting the prediction flow for both the LSTM and SVR model.

3.1. Data Acquisition

The experimental data that was used in this project is from the National Aeronautics and Space Administration (NASA) lithium-ion battery charge and discharge experimental data set.

The battery numbers in the dataset are, B7, B6, B5 and they all have a rated capacity of 2 Ah. The charge and discharge experiments were all carried out at room temperature of 24°C. The charging experiment was initially charged until the voltage reached 4.2 V with a constant current of 1.5 A, and then in constant voltage mode until the current dropped to 20 mA. The discharge process was initialized with a constant current of 2 A until the voltage dropped to 2.2 V (B0007), 2.7 V (B0005) and 2.5 V (B0006) respectively. The charging process and the discharging process in each charging and discharging cycle started from time 0 second in which the voltage, current, temperature and actual capacity were recorded. The experiments stopped after the actual capacity was less than the Acceptable Performance Ratio capacity of 70%. The battery capacity was recorded at the end of each discharge cycle. Therefore, this paper focuses on the discharge process and uses it to predict and estimate the SOH of the batteries.

3.2. Data Processing

Data preprocessing is a key stage in the machine learning process because machine learning algorithms can only function as best as the quality of data fed into them. Therefore, it is a very essential and crucial stage in the prediction process

In the dataset selected for the learning process, the battery capacity was recorded at the end of each discharge cycle due to the cyclic nature of the experiments. Therefore, the dataset was filtered into discharged cycles. Each cycle starts at time 0 seconds until the voltage dropped to 2.2 V (B0007), 2.7 V (B0005), 2.5 V (B0006). From the data extracted, the actual capacity degradation curve per discharge cycle of the batteries is shown in **Figure 2**.

At end of the data preprocessing exercise, the total number of cycles for each battery in the NASA dataset is shown in **Table 1** below.

Due to the heterogenous conditions under which the experiments for each battery took place. The processed data of each battery was fed into each of the models separately and the results were analyzed individually.

3.3. Feature Selection

Initial observation during the data preprocessing stage indicated a similar downward trend in the capacity of the batteries as discharge cycles increase for each battery dataset as shown in **Figure 2**.

To further visualize the correlation between the data and the actual capacity degradation, the discharging voltage timing curve, and discharging temperature timing curve of each battery were plotted in **Figure 3** and **Figure 4** respectively.

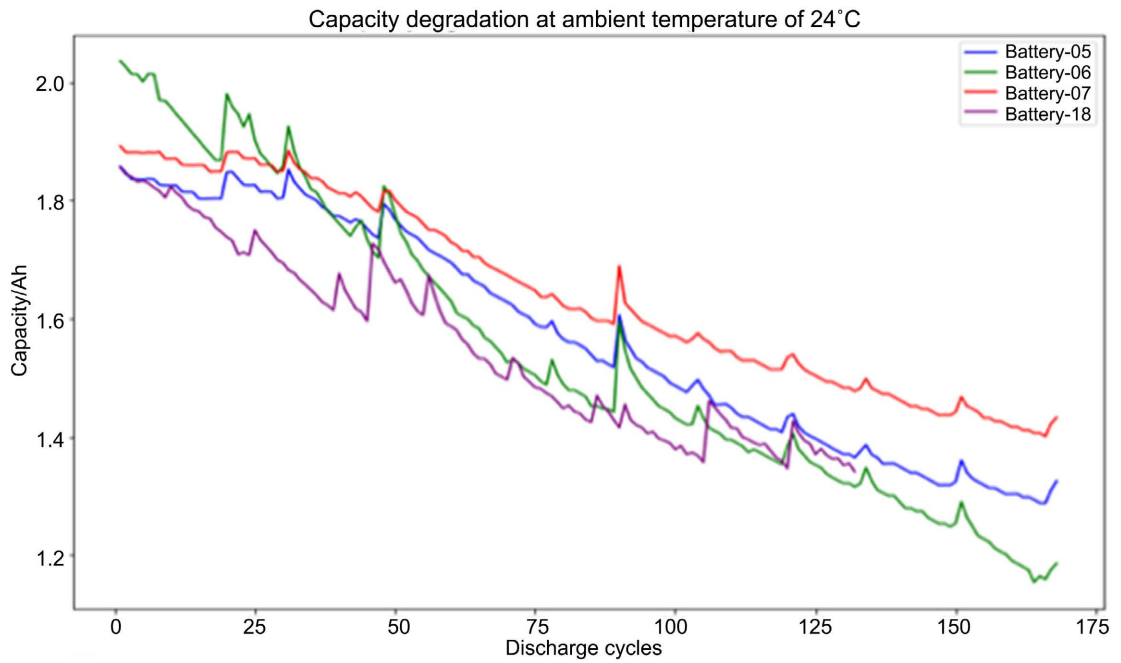


Figure 2. NASA lithium-ion battery discharge degradation curve.

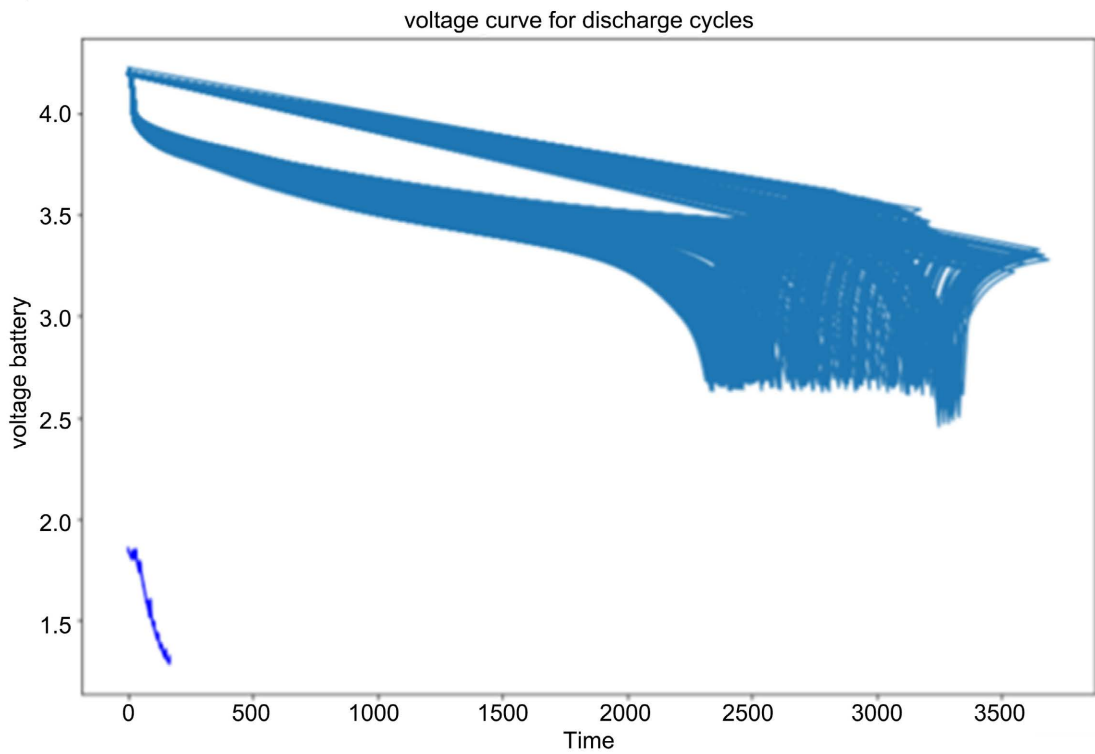


Figure 3. Discharge voltage trend.

It can be observed from **Figure 3**. That the time required for the voltage of different discharging cycles to reach 4.2V is different. It was also observed in **Figure 4** that the time it takes for the temperature of different discharging cycles to reach its maximum value increases as the discharging cycle increase.

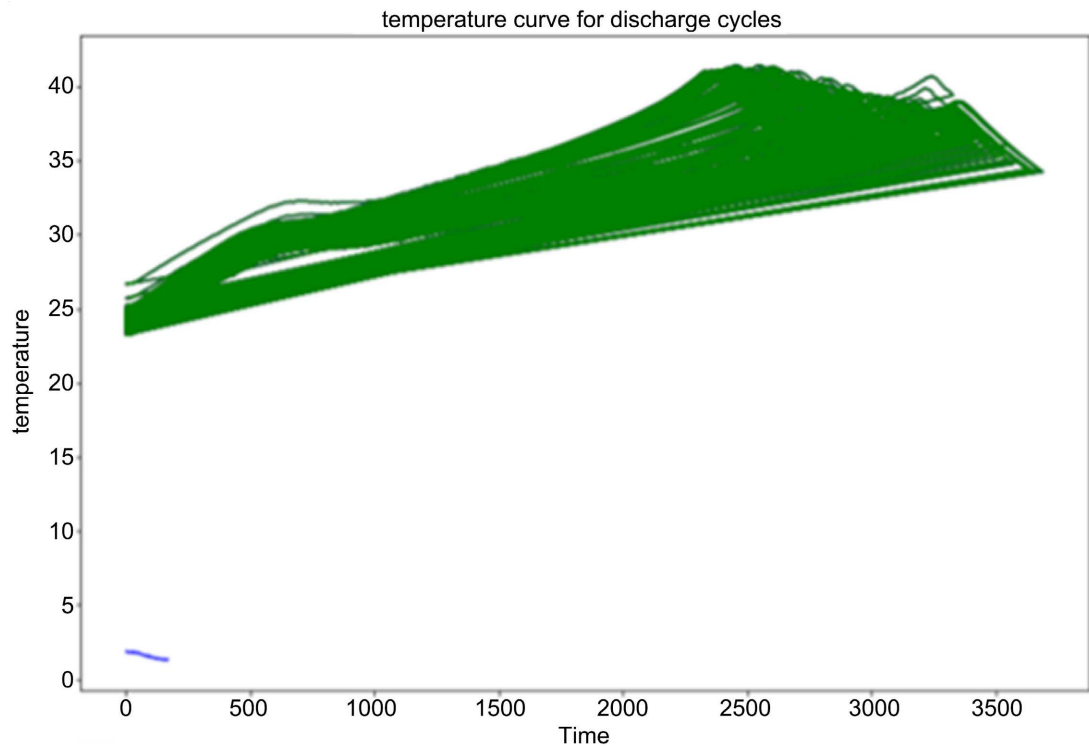


Figure 4. Discharge temperature trend.

Table 1. Discharge cycles.

Battery	Cycles
B5	168
B6	168
B7	168

To further shed more light on which features to be chosen, a feature heat correlation map of the dataset as shown in **Figure 5** was prepared. The heat map showed a high correlation between the discharge time, temperature and discharge cycles.

Therefore, due to the trends observed above; the number of discharge cycles, the time required for the voltage of different discharging cycles to reach 4.2 V as well as the time required for the temperature of different discharging cycles to reach a maximum value were the features extracted to form the dataset used in the prediction process.

3.4. Training and Prediction

There are several machine learning and deep learning models such as Logistic Regression, Support Vector Machine, Decision Tree Classifier, and Random Forest Classifier, RNN, LSTM, ANN with each of these models having inherent advantages and disadvantages depending on the domain they are being applied on.

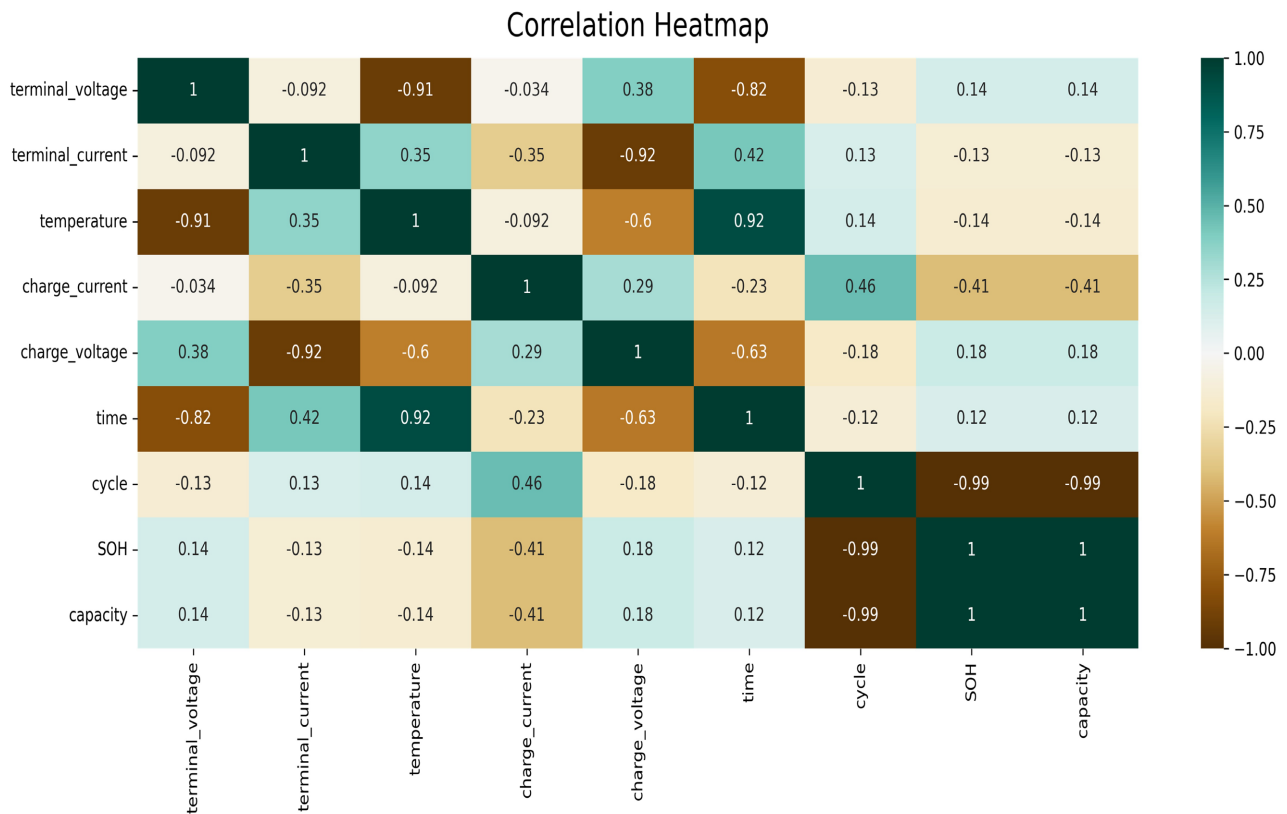


Figure 5. Feature correlation HeatMap.

However, the SOH estimation is a non-linear regression problem and therefore a robust ML model such as SVR was chosen due to its nonlinear kernel feature. The deep learning LSTM model was also considered in this paper because LSTM units include a memory cell that can maintain information in memory for long periods of time. Therefore, due to their memory ability, LSTMs in time series perform well on datasets in which new data points are highly dependent on historical data as with the case of the battery dataset. These inherent features of the models chosen should help the model fit the outliers observed in data points and the general non-linearity of the dataset. In this research, training, and testing of the model were performed using SVR and LSTM models. The optimal model was chosen based on the Mean Square Error (MSE) and Coefficient of determination (COD) performance metrics.

1) LSTM: LSTM was first proposed by Sepp Hochreiter and Jürgen Schmidhuber, and its framework was built from the RNN model for time series processing. Time series hold long-term memory information. However, for functional applications of the RNN model, it is challenging to deal with long-term dependence due to the gradient disappearance and gradient explosion in the algorithm. Due to the loss of long-term information, the analysis performance of a time series is very limited. The difference between LSTM and traditional recurrent neural networks is that LSTM adds a processor in the algorithm, which can judge whether the information is useful or not. The structure of the processor is

named as a memory cell. There are three gates inside of a memory cell; input gate, forget gate, and output gate, as shown in **Figure 5**. Using learned rules, messages that enter the LSTM network are judged. Only information that complies with the algorithm's certification will be maintained, and information mismatched will be overlooked through the forget gate. In addition, each cell contains several neurons. The autoregressive connection weight will remain at 1.0, ensuring that the state of the cell will remain constant as the time step changes without any external interference. **Figure 6** describes how the memory cells of the LSTM are updated at each time t .

In the training of neural networks, insufficient data leads to overfitting problems, which refer to the phenomenon for which a model has a fixed memory of the training data. This makes the output performance of the training data set excellent, while the performance of the validation dataset is extremely poor. To solve overfitting problems, the dropout method was proposed [12]. The core idea of dropout is optimizing the network to be thinner by integrating all the subnetworks via the removal of non-output units from the primary network and reducing the computation burden with the same training parameters [12]. In this research, due to the cyclic nature of the experiments resulting and the small dataset available for the training of the deep learning model. Dropout was applied to avoid overfitting the training dataset.

Each battery dataset was divided into training and validation/testing data. The data designated for training the model were fed into the LSTM model using the time series look back approach. This approach was implemented because we wanted to observe how LSTM uses its memory capabilities in predicting the future trend of the battery capacity as the discharge cycle increases. We also wanted to observe the model's performance especially as regards predicting outlier data points caused by sudden upward spikes in the capacity as it descends per discharge cycle. This was studied to be caused by the internal chemical composition of the battery. Therefore, the LSTM implementation using the look-back mechanism was constructed and tuned to properly exploit the memory abilities of the LSTM model. The model was trained with previous data in specified steps

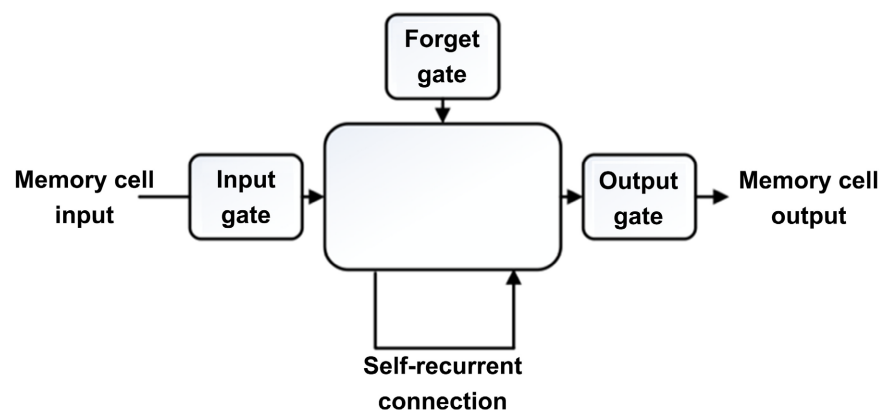


Figure 6. Structure of a memory cell.

in form of a time series and used to predict a specified number of steps of the output.

The LSTM model was fitted with multiple hidden layers using several neuron values as well as the dropout layers in order to get the best result possible and avoid overfitting (Figure 7).

The optimal parameters as well as prediction curves of the battery dataset will be highlighted and discussed in the results section.

2) SVR: Support vector machine was developed by Cortes and Vapnik, based on the structural risk minimization Statistical Learning Theory, and the Vapnik-Chervonenkis (VC) theory. In machine learning, Support Vector Machines is a type of supervised learning model with inherent learning algorithms that analyze data used for classification and regression analysis. Therefore, Support Vector Machine can also be used as a regression method while maintaining all the main features such as maximal margin that characterize the algorithm. In the regression method, because the output is a real number, it becomes very difficult to predict the information at hand due to its infinite possibilities. Therefore, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already been requested from the problem. Given data points, the SVR tries to find the best curve that fits all data points. However, the regression algorithm uses the curve to find the match between the vector and position of the curve instead of using the curve as a decision boundary as in the classification case.

Support Vectors help in determining the closest match between the data points and the function which is used to represent them. SVR also contains the kernel feature that makes it possible to perform linear and non-linear analysis. During non-linear analysis, the Kernel converts the data into a higher dimensional feature space to make it possible to perform the linear separation and improve the generalization ability, and finally get the non-linear learning algorithm in the initial low-dimensional space.

The grid search method shown in Figure 8 was applied during the training in order to obtain the optimal hyperparameters for the SVR model.

```
model: sequential_0/
```

Layer (type)	Output Shape	Param #
lstm_296 (LSTM)	(None, 5, 100)	42000
dropout_296 (Dropout)	(None, 5, 100)	0
lstm_297 (LSTM)	(None, 5, 100)	80400
dropout_297 (Dropout)	(None, 5, 100)	0
lstm_298 (LSTM)	(None, 5, 100)	80400
dropout_298 (Dropout)	(None, 5, 100)	0
lstm_299 (LSTM)	(None, 5, 100)	80400

Figure 7. LSTM model information.

```

parameters = {'kernel': ('linear', 'rbf', 'poly'),
              'C': [1,2,3,4,5,6,7,8,9, 10], 'gamma': [0.00001,1e-7, 1e-4],
              'epsilon': [0.0001,0.1,0.2,0.5,0.3]}
svr = SVR()
clf = GridSearchCV(svr, parameters,scoring='neg_root_mean_squared_error')
clf.fit(X_train,y_train)
print(clf.best_params_)

```

Figure 8. GridSearch hyperparameter turning snippet.

The optimal parameters and metrics obtained during the SVR training and testing process are discussed in the next section.

4. Results and Discussion

The implementation of the two models was programmed using the python language and the Jupyter notebook IDE. The python files for each model were compiled on the LEAP cluster by submitting SLURM job files. The Mean square error and coefficient of determination were the performance metrics used as a guide in determining the optimal results during the training validation and testing process. These metrics were also used as the basis of comparison to determine the best model.

Table 2 compares the MSE of both models for all the batteries considered in this paper.

The obtained coefficient of determination for each model is listed in **Table 3**.

To obtain optimal results during the SOH estimation process for the LSTM model, the first half of preprocessed data were used to train the LSTM network model. The remaining data were used to examine the accuracy of model predictions.

The optimal LSTM prediction regression curves displaying the actual and predicted data for the B5, B6 and B7 datasets are shown in **Figures 9-11** respectively.

During the SOH estimation process for the SVR model, 80% of the dataset was used to train the model. The remaining data were selected to test the accuracy of model predictions.

The grid search tool was utilized in obtaining the optimal parameters and this process returned the SVR-RBF as the appropriate kernel to get the least mean square error and highest coefficient of determination. The optimal prediction regression curves for the SVR model displaying the actual and predicted data for the B5, B6, and B7 battery datasets are shown in **Figures 12-14** respectively.

It can be observed from the results of both models in **Table 2** and **Table 3** that the SVR performed better in the SOH estimation of the batteries dataset used in this paper. The SVR had the least mean square error which could also be validated by how close the predicted data in **Figures 11-13** conformed with the test

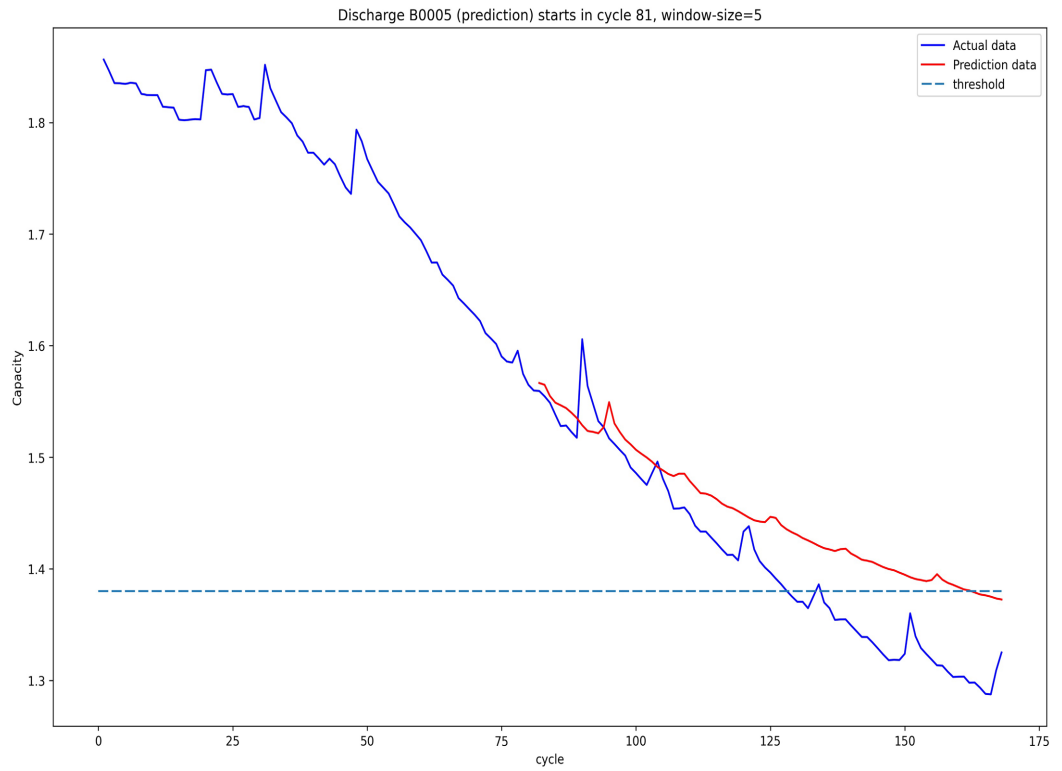


Figure 9. Optimal B5 LSTM model.

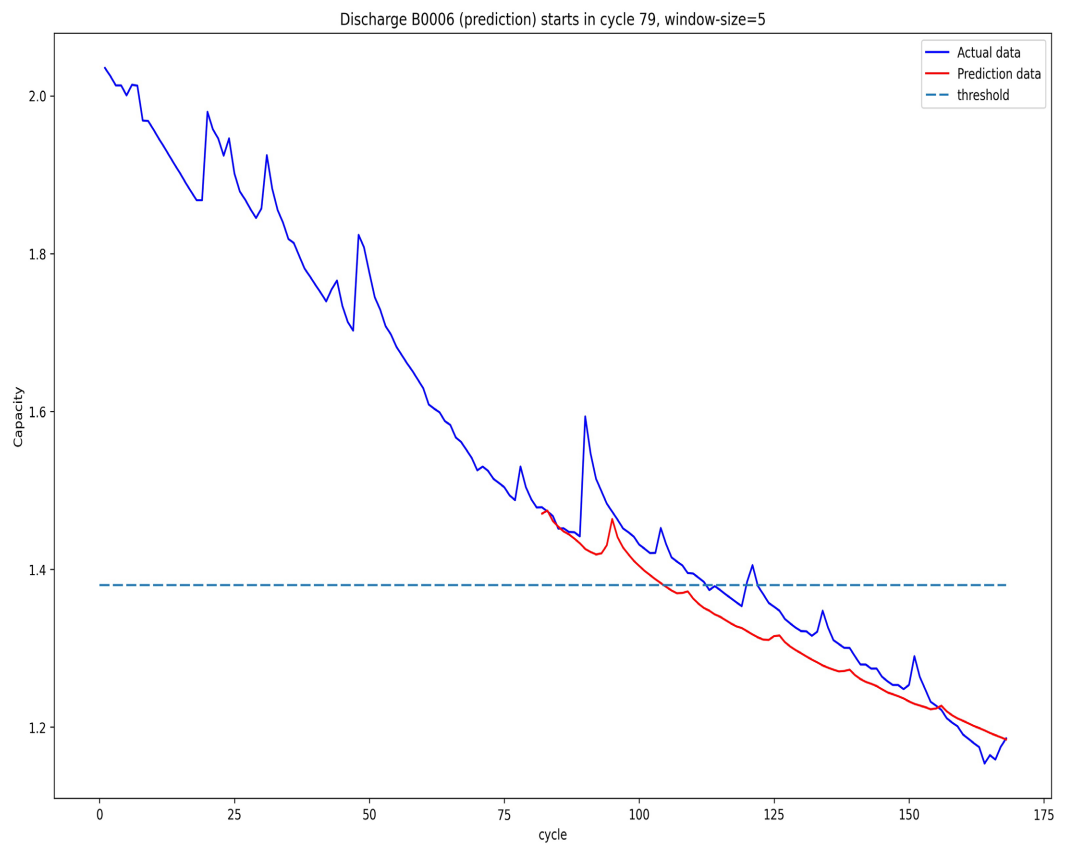


Figure 10. Optimal B6 LSTM regression plot.

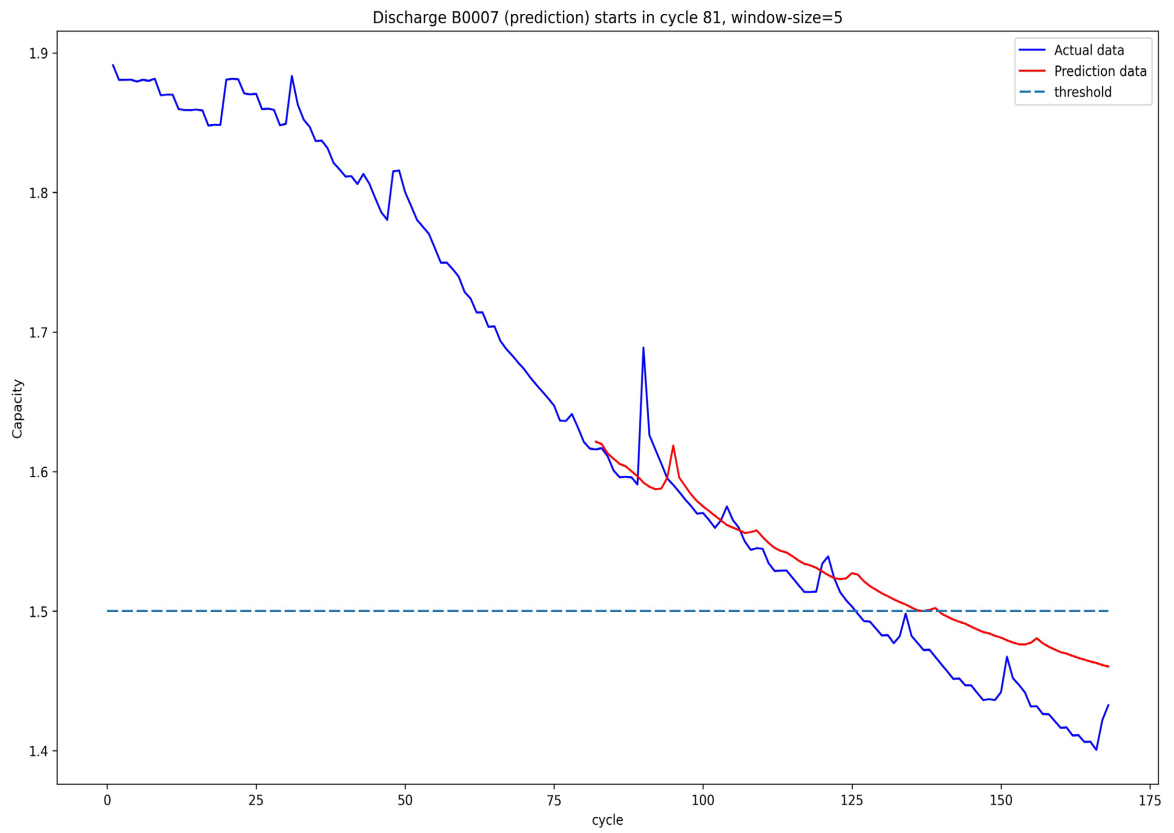


Figure 11. Optimal B7 LSTM model regression plot.

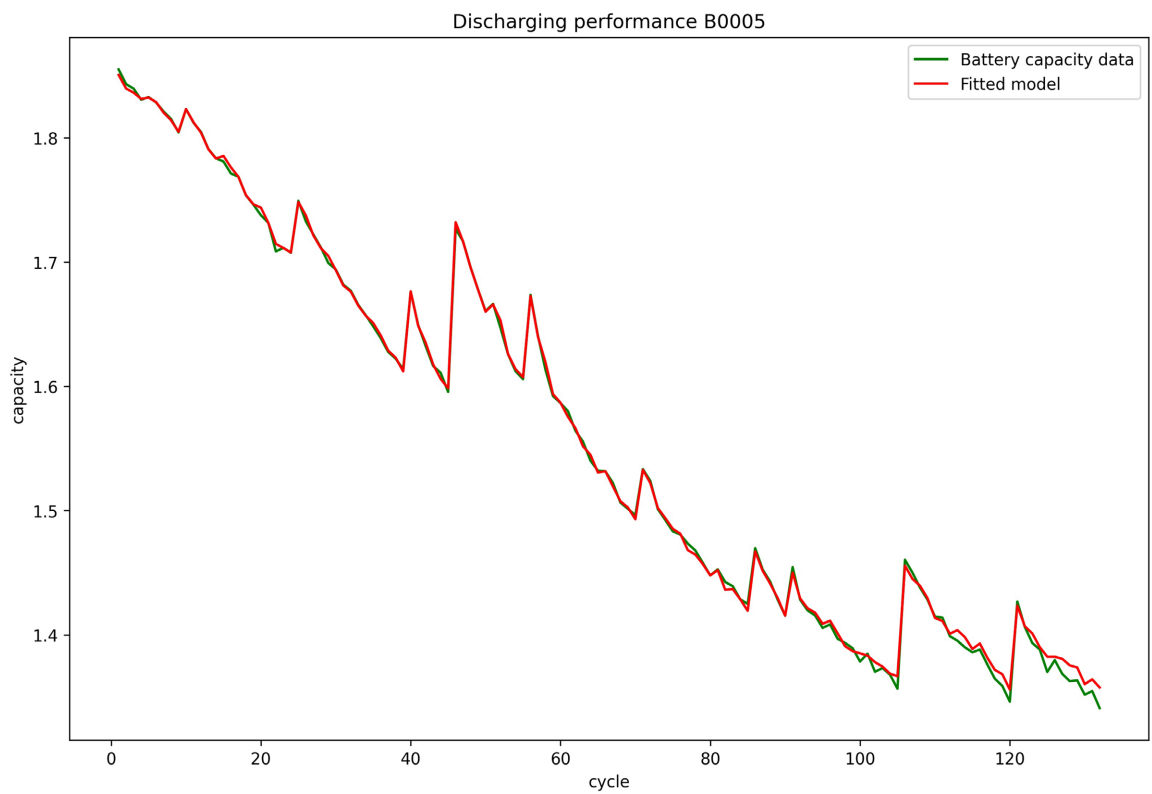


Figure 12. Optimal B5 SVR model regression plot.

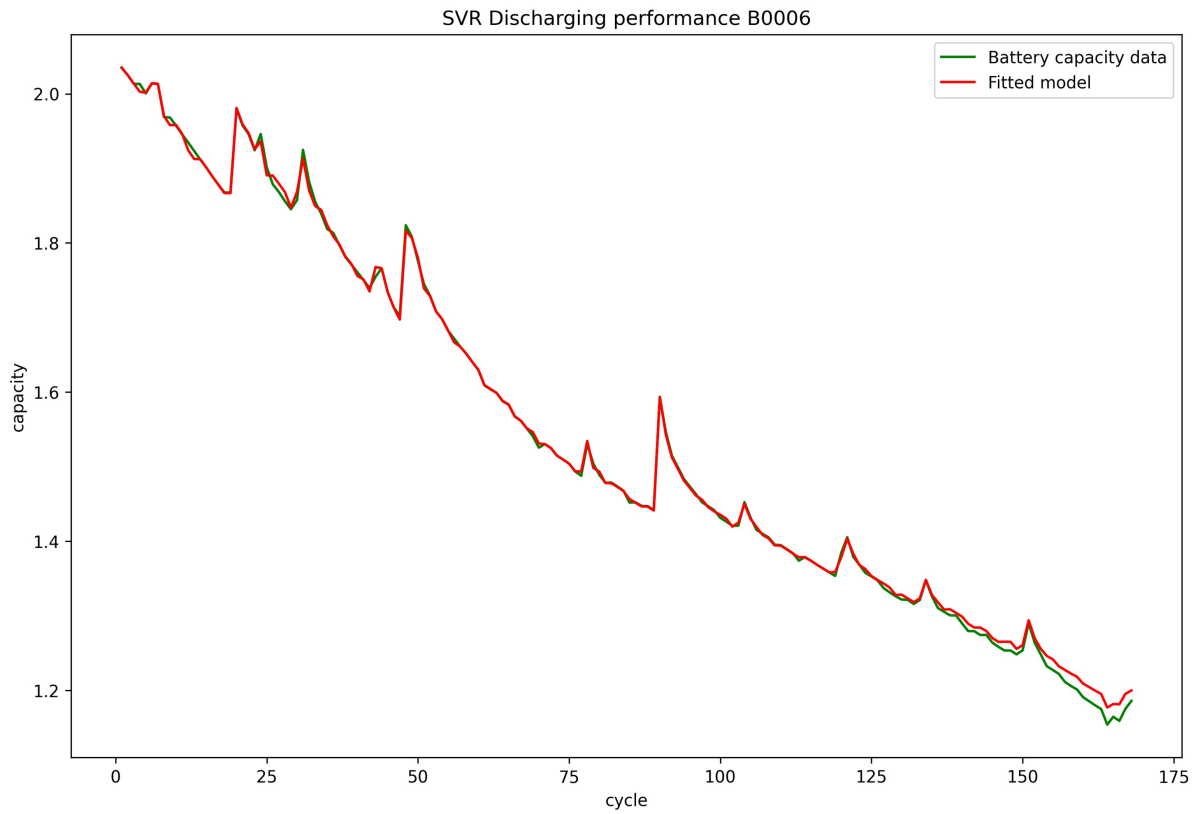


Figure 13. Optimal B6 SVR model regression plot.

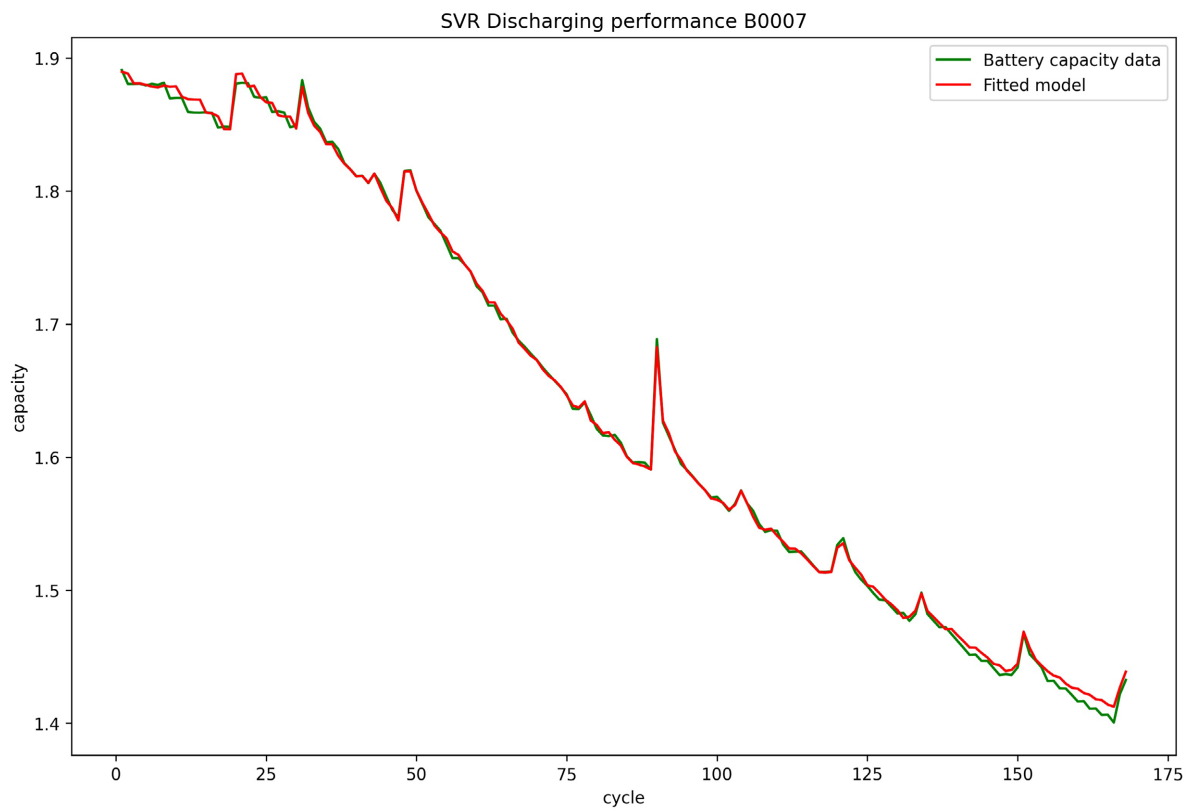


Figure 14. Optimal B7 SVR model.

Table 2. MSE of each battery.

	B5	B6	B7
LSTM	0.034	0.041	0.039
SVR	0.003	0.013	0.016

Table 3. Cod of each battery.

	B5	B6	B7
LSTM	0.84	0.83	0.67
SVR	0.98	0.93	0.92

Table 4. Optimal hyperparameters.

	C	Gamma	Kernel	Epsilon
SVR	10	0.0001	RBF	168

data. Therefore, based on the performance metrics SVM-RBF is the best classifier for the SOH estimation of Lithium-ion batteries. The Grid search was performed to obtain the optimal hyperparameters for the SVR. The optimal hyperparameters for the SVR are listed in **Table 4**.

The SVR model was also able to successfully learn and to fit the several outlier data points in the dataset. Therefore, validating the theory that SVR's inherent ability to model non-linear data by converting it to high dimensional hyper-planes using the RBF kernel was suitable in the SOH estimation and could also excellently predict SOH despite several outliers recorded during the battery lifetime.

The LSTM model using the time series look back strategy suffered from overfitting towards the end of each battery dataset considered. Despite the introduction of several dropout layers as well finely controlling the batch size during training. Another downside to the time series approach was observed through the course of the LSTM learning process. The model exhibited a noticeable delay especially at points when the model was trying to predict the sharp and sudden capacity upward trends that occurred at random intervals throughout the discharge lifecycle of the batteries considered in this paper.

5. Conclusions

In this paper, the importance of lithium-ion batteries and their usage were introduced. The industry definition of battery SOH and APT were further explained. The data acquisition and preprocessing sections introduced the experimental conditions of the NASA lithium-ion battery under which the dataset was obtained as well as the data cleaning process. The various methods of feature extractions, the training, and validation and the results of the learning process were further discussed in the succeeding sections.

At the end of the learning process, the results of the LSTM and SVR models were compared using the mean square errors and the coefficient of determination as the performance metrics. It was observed that the SVR model performed better and is the optimal model of the two learning approaches considered in this paper. This indicates that the novel approach of tackling the prediction as a regression issue yielded the overall best model of predicting the state of health of the batteries.

Conflicts of Interest

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

References

- [1] Garche, J. and Jossen, A. (2016, August) Monitoring and Safety Tests of Batteries: From State of Charge (SOC) and Health (SOH) to Misuse, Abuse and Crash. In: *AIP Conference Proceedings*, Vol. 1765, Article ID: 020005, AIP Publishing LLC. <https://doi.org/10.1063/1.4961897>
- [2] Jiang, L. (2013) Research on Lithium Ion Battery SOC Estimation and RUL Prediction. University of Electronic Science and Technology of China, Chengdu, 4-5.
- [3] Saha, B. and Goebel, K. (2007) Battery Data Set, NASA Ames Prognostics Data Repository. NASA Ames Research Center, Moffett Field, CA. <http://ti.arc.nasa.gov/project/prognostic-data-repository>
- [4] Weicker, P. (2013) A Systems Approach to Lithium-Ion Battery Management. Artech House Publisher, Boston.
- [5] Qing, D., Huang, J. and Sun, W. (2014) SOH Estimation of Lithium-Ion Batteries for Electric Vehicles. 2014 *Proceedings of the 31st ISARC*, Sydney, 925-928. <https://doi.org/10.22260/ISARC2014/0125>
- [6] Pózna, A.I., Magyar, A. and Hangos, K.M. (2017) Model Identification and Parameter Estimation of Lithium Ion Batteries for Diagnostic Purposes. 2017 *International Symposium on Power Electronics (Ee)*, Novi Sad, 19-21 October 2017, 1-6.
- [7] Lin, C., Tang, A. and Wang, W. (2015) A Review of SOH Estimation Methods in Lithium-Ion Batteries for Electric Vehicle Applications. *Energy Procedia*, **75**, 1920-1925. <https://doi.org/10.1016/j.egypro.2015.07.199>
- [8] Zou, Y., Hu, X., Ma, H. and Li, S.E. (2015) Combined State of Charge and State of Health Estimation over Lithium-Ion Battery Cell Cycle Lifespan for Electric Vehicles. *Journal of Power Sources*, **273**, 793-803. <https://doi.org/10.1016/j.jpowsour.2014.09.146>
- [9] Moura, S.J., Chaturvedi, N.A. and Krstić, M. (2014) Adaptive PDE Observer for Battery SOC/SOH Estimation via an Electrochemical Model. *Journal of Dynamic Systems, Measurement, and Control*, **136**, Article ID: 011015. <https://doi.org/10.1115/1.4024801>
- [10] Chen, L., Lv, Z., Lin, W., Li, J. and Pan, H. (2018) A New State-of-Health Estimation Method for Lithium-Ion Batteries through the Intrinsic Relationship between Ohmic Internal Resistance and Capacity. *Measurement*, **116**, 586-595. <https://doi.org/10.1016/j.measurement.2017.11.016>
- [11] Berecibar, M., Gandiaga, I., Villarreal, I., Omar, N., Van Mierlo, J. and Van Den Bossche, P. (2016) Critical Review of State of Health Estimation Methods of Li-Ion

Batteries for Real Applications. *Renewable and Sustainable Energy Reviews*, **56**, 572-587. <https://doi.org/10.1016/j.rser.2015.11.042>

- [12] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, **15**, 1929-1958.

Abbreviations and Acronyms

SOH—State of Health
LSTM—Long Short-Term Memory
RNN—Recurrent Neural Networks
SVR—Support Vector Regression
SVM—Support Vector Machines
COD—Coefficient of Determination
BMS—Battery Management System
SOC—State of Charge
MSE—Mean Square Error
ANN—Artificial Neural Network
EV—Electric Vehicles