Machine Learning Integration with Battery Management System

Navale Shubham Dagadu[1], Divase Pranita Satish[2], Gopale Sanket Lahanu[3],

Ghungrale Aarti Gangadhar[4], Prof.M.D.Landge[5]

[1,2,3,4,5] Department of Electrical Engineering

AmrutvahiniCollege of Engineering, Sangamner, Maharashtra, India

Abstract: *The durability and reliability of battery management systems in electric vehicles to forecast the state of charge (SoC) is a tedious task. As the process of battery degradation is usually non-linear, it is extremely cumbersome work to predict SoC estimation with substantially less degradation. This paper presents the SoC estimation of lithium-ion battery systems using six machine learning algorithms for electric vehicles application. The employed algorithms are artificial neural network (ANN), support vector machine (SVM), linear regression (LR), Gaussian process regression (GPR), ensemble bagging (EBa), and ensemble boosting (EBo). Error analysis of the model is carried out to optimize the battery's performance parameter. Finally, all six algorithms are compared using performance indices. ANN and GPR are found to be the best methods based on MSE and RMSE of (0.0004, 0.00170) and (0.023, 0.04118), respectively.*

Keywords: lithium-ion battery; battery management; sustainable energy; machine learning algorithms; electric vehicles; state of charge

I. INTRODUCTION

The transport industry accounts for the bulk of greenhouse gas emissions and pol- lution to the environment [1]. The transport sector can be improved by the introduction of the e-mobility applications such as electric vehicles (EVs) [2], hybrid locomotives and other battery-energy storage systems [3]. The energy storage system is one of the most significant parts of EVs and smart grid technologies [4–7]. The smart grid technology is the emerging technology in electricity transmission and distribution lines. Numerous batteries are available in the market for various energy storage applications. Specifically, lithium-ion batteries are selected as an energy storage technology for EVs due to its gravimetric and volumetric density, high hour's efficiency, and long life [8,9]. However, thermal manage- ment of batteries for EV application is important [10]. EV charging stations are widely used internationally, and ports have been expanded at public and private charging points [11]. In Belgium, two EVs with different battery capacities are investigated. It is reported that the grid utility for the EVs leads to volatility in power supply, electricity quality and grid control issues [12,13]. Currently, research is going on to transform buildings from energy consumers to energy producers by integrating renewable energy systems into the building where storage systems can play an indispensable role [14–20]. Photovoltaic energy can be stored in batteries instead of pumped into the grid [21,22]. Presence of battery energy storage can also enable integrating the electrically activated smart window integration into buildings [23,24].

The development of data-driven algorithms such as machine learning methods has taken a major step in recent years to improve the accuracy of state of charge (SoC) mea- surements with improved generalization performance, improved learning capabilities for high accuracy and convergence [25]. The battery's performance is estimated based on state of health (SOH) [26] and remaining useful life (RUL). The SOH and RUL of the battery can be predicted by artificial intelligence such as machine learning and the deep learning approach which are intelligent and adaptive. However, estimation and prediction outcomes are subjected to the collection of trained data. The artificial neural network (ANN) approach has an exceptional potential to construct a non-linear map between input and output parameters which illustrates the non-linear model complexity. However, the accuracy of the data-driven approach depends on the quantity and quality of the data; which may lead to problems with the over-fitting and under-fitting of the data. The ANN method can form a non-linear map to show a complicated non-linear model. Dual ANNs enable modelling of an open-circuit voltage (OCV)-based approach to estimate SoC [27]. The linear ANN battery model is used to define the first or

second-order parameters of the electro-chemical model. The SOH and RUL batteries with machine learning methods are expected to develop big data and the artificial intelligence industry. The RUL predic- tion method has been developed in a model [28] consisting of relevance vector machine (RVM), unscented Kalman filter (UKF) and a full empirical decomposition ensemble. The RVM is employed to change the impacts of the UKF by forecasting the SoC. On the other hand, support vector machines (SVM) are also a well-known machine learning method for SOH and RUL estimation [29]. A new way of estimating SOH using an established probabilistic knowledge-based neural network (PKNN) and Markov chain to address the uncertain external condition and the dynamic internal electrochemical mechanism is proposed. When matching partial loads [30] the predictive diagnostic method is de- veloped [31]. The Gaussian process regression (GPR) is an approach that can deal with complexity in a complex model similar to a Bayesian non-parametric approach [32]. Auto- matic GPR is used to capture control, temperature and SoC mapping, considering the test of electrochemical information for covariance functions of GPR [33]. The Long short-term memory(LSTM) model is used for estimating the residual GPR [25] and SoH of battery supply approaches [34] in GPR system. In comparison, the second back propagation of the NN (BPNN) caught the relationship between OCV and SoC [35]. It is reported that the radial base function neural network focused on ambiguity (RBFNN). RSAM algorithm was developed to simulate bias feature for the multi-cell battery pack SoC estimation [36]. In practice an ANN model is developed to predict the SoC estimation based on the load classification which includes post-processing and boost overfitting [37]. The long short term memory recurrent neural network (LSTM-RNN)-based SoC estimation for the lithium- ion batteries is also present [38] as well as the technique of BPNN for estimation of SoC forecasts has been investigated before [39,40]. A new design based on a deep feedforward neural network is used for the prediction of SoC estimation of the battery. The solution suggested that predicted SoC converged easily, although the real-time SoCs predicted are wrong. The inclusive analogy circuit model is the non-linear radial basis function neural network (RBFNN) algorithms to predict the approximate SoC value is developed [41]. An improved non-linear autoregressive network with exogenous inputs neural network (NARXNN) algorithm is developed to estimate battery SoC [42]. The abating hysteresis technique is used to detect the right value for input delays, feedback delays and hidden layer neurons [43]. The multi-layer Levenberg–Marquardt (L–M) wavelet neural network model (WNN) is developed to optimize the battery performance. By using the particle swarm optimization (PSO) algorithm, the SoC approximation is investigated. The recurrent neural network is developed to estimate the SoC estimation of the battery [44]. A structural heating, ventilation air conditioning (HVAC) analysis of electric vehicles, home appliances, distributed power generation and electrical storage are being carried out based on the artificial intelligence of the system to estimate SoC [45–47]. Recent studies on battery charging, discharging characteristic and state of estimation studies are listed in Table 1. The SVM algorithm is also used to process Application programming interface (API) weather data [48]. Results demonstrate a good forecast for photovoltaic panels to optimize energy production and cargo balance. The distribution grid defines optimization algorithms for the assignment and operation of stationary and EV batteries. The results suggest that a significant drop in battery size and a minimal energy loss can be accomplished by EV sim- ulation and prediction of load [49] and Photovoltaic(PV) interactions [13,50–52]. Besides, data-driven methods and battery performance evaluations rely not just on the choice of health metrics but also on the battery model range. The state of charge (SoC) estimation of the battery is one of the important parameters for the battery management system (BMS) in electric vehicles. (current, voltage) average, area) voltage curve voltage curve charging curve discharging curve curves (peak, valley)

In the present work, the SoC of lithium-ion batteries is predicted based on the six machine learning algorithms using data derived from the electric vehicle BMS. The algo- rithms used in the studies are artificial neural network (ANN), support vector machine (SVM), linear regression (LR), Gaussian process regression (GPR), and ensemble bagging (EBa), and ensemble boosting (EBo) algorithm. Finally, all six algorithms are compared with performance indices.

Table 1. Earlier studies on the battery charging and discharging characteristics in machine learning.

Feature Parameter	Battery	Performance Index and Precision		Reference
The energy of the signal (current, voltage)	NASA 18650	MAE	<1.29%	[53]
Temperature (min, max, average, area)	NASA 18650	RMSE	<3.58%	[54]
The slope of the charging voltage curve	NASA 18650	RMSE	<3.45%	[55]
The slope of the discharging voltage curve	NASA 18650	RMSE	<3.84%	[56]
Equal voltage drops in charging curve	NCM/ graphite	RMSE	2%	[57]
Equal voltage drops in discharging curve	NASA 18650	MAE	<1.29%	[58]
The characteristic of I.C. curves (peak, valley)	Prismatic Li-ion Battery	RMSE	2.99%	[59]

II. MATERIALS AND METHOD

The Panasonic 18650FP battery cell (Panasonic, Zellik, Belgium) is used for experimen- tal data sets for an electric vehicle. The research equipment used to compile experimental data sets includes the tested batteries, host computer, battery programming, battery dis- charge, thermal chamber, stress, current, temperature and electrical quantity instruments. Panasonic's 18650PF dataset [38] is compiled at McMaster University, Ontario, Canada, by the Department of Mechanical Engineering. The battery datasets acquired are trained, validated and tested using MATLAB version 2020b (MathWorks, Natick, MA 01760, USA) which is carried out on a 24 GB Quadro NVIDIA RTX 6000 workstation computer with an Intel i9 processor. MATLAB 2020b's Neural Network Toolbox, Regression Toolbox and Statistics and Fitting Toolbox are the toolboxes used in this experiment.
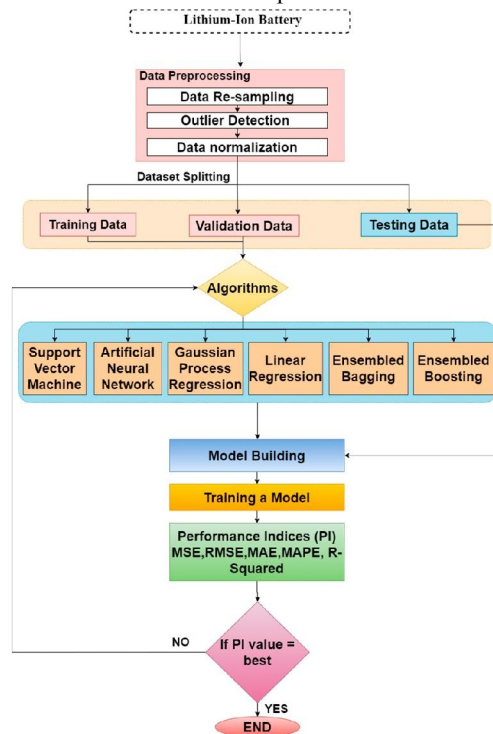

Figure 1. Overall flow chart of the proposed state of charge estimation of the battery Mean Square error (MSE), Root Mean Square error (RMSE), Mean Absolute Error (MAE) and Mean Absolute percentage Error (MAPE).

First, the suggested machine learning (ML) algorithm is used according to known partial data to construct a predictive model of the state of charge (SoC) for forecasting the complete charging curve. The flow chart of the proposed SoC method is illustrated in Figure 1. The overall SoC diagnostics framework is based on short-term charging results.

The method suggested consists of three modules: input parameters, feature extraction and machine learning (ML) algorithms, and estimates the SoC. Six different algorithms are adopted in this study, namely, ANN, SVM, LR, GPR, ensemble bagging and ensemble boosting algorithms.

2.1. Batteries State of Charge Estimation

The machine-learning (ML) algorithms are specifically used to create an accurate SoC. The following are all the main sections shown in the flowchart. The proposed SoC estimation method is validated under a wide range of battery operating conditions. The ML output data are then removed, and the SoC features are collected. Finally, the non-linear cartography between input and SoC functions provides SoC diagnosis with the aid of a serviceable model for all different algorithms such as ANN, SVM, LR, GPR, ensemble bagging and ensemble boosting. The training module ANN, SVM, linear regression, GPR, ensemble bagging and ensemble boosting comprises the critical parameter optimization process. The ML algorithm-based estimation of the SoC can be predicted from the four essential parameters of a lithium-ion battery such as battery current, battery voltage, battery capacity and temperature of the battery based on the available dataset. The ML algorithm used in this study is elaborated in the following sections.

2.2. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are parallel processing approaches that can specif- ically describe non-linear and complex interactions using input-output data set training patterns. ANNs provide non-linear mapping between inputs and outputs through intrinsic capacity. The ANNs' ability to learn system behavior from the representative data enables ANNs to solve numerous complex large-scale problems. The ANN algorithm follows, which is shown in Figure 2.
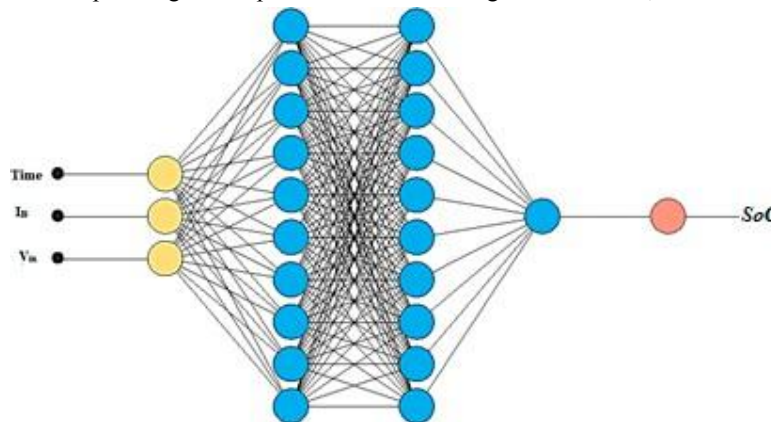

Figure 2. Input and output layer of artificial neural network.

Step 1: Randomly initialize the weights and bias of the model

Step 2: Log sigmoid activation function is used in the hidden layer

$$Sig(net) = 1/1 + e(-net) \quad (1)$$

For input variable m, the j-th input layer node holds $x_{m,j}$.

The overall input to the k-th node in the hidden layer is

$$net_k = \sum_{j=0}^{n} w_{k,j}\, x_{m,j} + \theta_{k,j} \quad (2)$$

where, $w_{k,j}$ = weight from the input layer to the hidden layer, $\theta_{k,j}$ = bias from the input layer to the hidden layer. The hidden layer output at l-th node is given by

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 5, May 2025

Impact Factor: 7.67

$$net_k = \sum_{j=0}^{n} w_{k,j} \, x_{m,j} + \theta_{k,j} \qquad (3)$$

The overall l-th node in the output layer is given by

$$x_{m,l} = S_h\left(\sum_{l=0}^{n} w_{k,j} x_{m,l} + \theta_{k,j}\right) \qquad (4)$$

wl,i= weight from the hidden layer to the output layer, θl,j = bias from the hidden layer to the output layer.
The final output layer is given by

$$SoC_{l,m} = S_o\left(\sum_{l} w_{l,k} x_{m,l} + \theta_{l,k}\right) \qquad (5)$$

where SoCl,m is the estimated SoC and So is the activation function.
Step 3: The error estimated is backpropagated to the hidden layer from the output layer

$$\in o = So(1 - So)(SoCo - So) \qquad (6)$$

The hidden layer error is calculated by

$$\in h = Sh(1 - Sh) \in o \, wl,i \qquad (7)$$

Step 4: The weights and biases are updated using the weight equations.

2.3. Support Vector Machine (SVM)

The support vector machine (SVM) is a popular and commonly used soft computing technique in many areas. The basic principle of SVMs is to use non-linear mapping for data mapping in some areas and apply the linear algorithm in the function space. One form of SVM is the support vector regressor, which has been developed for regression problems. The SVM algorithm model is followed, which is shown in Figure 3. An empirical equation of proposed algorithm is presented in Table 2.
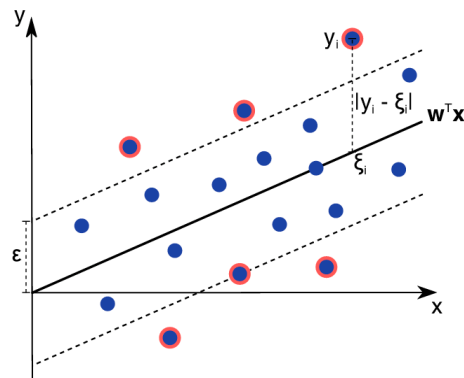

Figure 3. Support vector machine illustration of a function of SVM regression represented by wTx.

Algorithms Model	Empirical Equation
Artificial Neural Network	$y_i = s(\sum_{j=1}^{N} W_{ij} x_j + b_i)$ W_{ij} = weight to neuron i from neuron j b_i = bias x_j = input vectors
Support Vector Machine	$Y_i = \sum_{i}^{N} W.K\,(x_{i}, x) + B$ Y_i = predicted output W = weights K = kernel trick (x_i, x) = support vectors B = bias
Gaussian Process Regression	Test the model with testing data calculate the performance metrics for the tested data $P((y_i \mid f(x_i),\, x_i) \sim N\,(y_i h(x)^T \beta + f(x)),\, \sigma^2)$ σ^2 = noise variance β = coefficient vector $f(x_i)$ = observtion x_i
Linear Regression	$y^{\hat{}} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$ b_i = bias value $X_{(1,2,3,4)}$ = input feature values
Ensemble Bagging	Output the bagging model: $\hat{f}_{bag}\,(x) = 1/K \sum_{i=1}^{k} \hat{f}^{\times i}(x)$
Ensemble Boosting	The output of boosting tree: $\hat{f}\,(x) = \sum_{k=1}^{K} \hat{f}^{\times i}(x)$

Table 2. Empirical equation of proposed algorithm.

Step 1. Import the input features

Step 2. Analyze the correlation and directivity of the data Step 3. Split the dataset into the train and validation test

Step 4. Choose the kernel function out of (linear, polynomial, sigmoid, radial basis) Step 5. Train the model with training data

Step 6. Evaluate the model performance Step 7. Test the model with testing data

Step 8. Calculate the performance metrics for the tested data

2.4. Linear Regression (LR)

Linear regression algorithms will be applied if the output is a continuous variable. In contrast, grading algorithms are applied when output is broken up into sections like pass/fail, good/average/bad, et cetera. We have different regression algorithms or clas- sifying behavior, the LR algorithm being the fundamental regression algorithm. The linear model algorithm follows, which is shown in Figure 4. An empirical equation of the proposed algorithm is presented in Table 2.
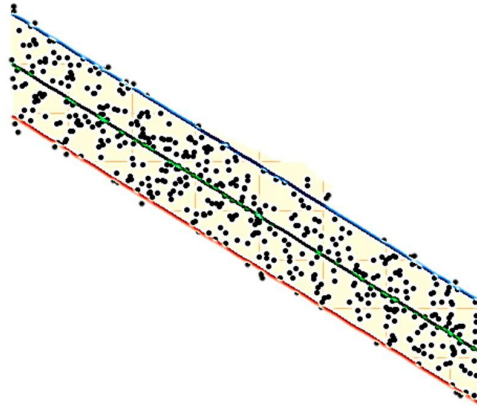
Figure 4. Linear regression.

Step 1. Get the input features

Step 2. Analyze the correlation and directivity of the data Step 3. Estimate the model

Step 4. Fit the best fitting line Step 5. Evaluate the model and

Step 6. Test the model with testing data

Step 7. Calculate the performance metrics for the tested data

2.5. Gaussian Process Regression (GPR)

Gaussian regression of process (GPR) is a non-parametric Bayesian regression method that generates waves in machine learning. GPR has various advantages, such as, it works well on small data sets and can provide predictive uncertainty measurements. The algo- rithm of the Gaussian process regression (GPR) model is shown in Figure 4. An empirical equation of the proposed algorithm is presented in Table 2.

Step 1. Import the input features

Step 2. Analyze the correlation and directivity of the data

Step 3. Split the dataset into the train and validation test

Step 4. Build the model for the Gaussian process regression model

Step 5. Train the model with training data

Step 6. Evaluate the model performance

2.6. Ensemble Bagging (EBa)

Bagging is a meta-algorithm machine-learning set designed to strengthen machine learning algorithms' accuracy and precision in statistical classification and regression. It also reduces variance and aids in over-fitting avoidance. Bagging is a way to reduce the uncertainty of estimation by producing additional data for dataset testing using duplication variations to generate different sets of initial data. Boosting is an iterative strategy that relies on the previous description to adjust the weight of the observation. The bagging trees algorithm is as follows, which is shown in Figure 5a.
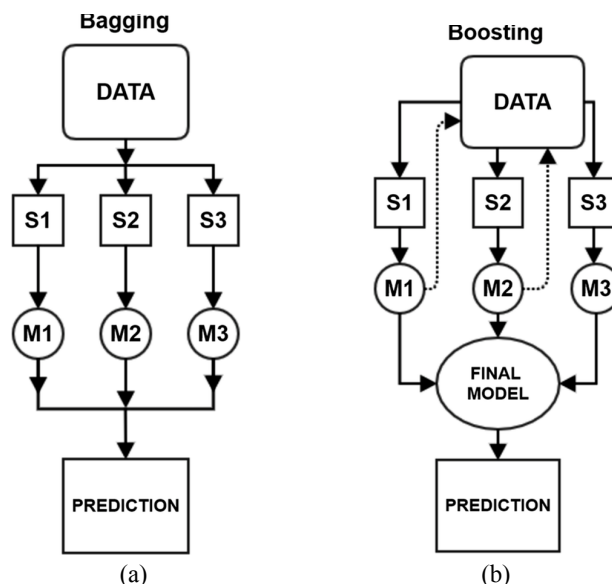
IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 5, May 2025

ISSN: 2581-9429

Impact Factor: 7.67

Figure 5. Pictorial representation of algorithms: (a) ensemble bagging and (b) ensemble boosting.

Step 1. for i = 1 to K, do

Step 2. Generate a bootstrap sample of the original data Step 3. Train an unpruned tree model on this sample Step 4. End

2.7. Ensemble Boosting (EBo)

Boosting is a whole sequential process that eliminates the bias error in general and generates good predictive models. The algorithm assigns weights to each resulting mode during training, shown in Figure 5b. An Algorithm Boosting Trees and empirical equation of the proposed algorithm is presented in Table 2.

Step 1. Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set

Step 2. Compute the average response, y, and use this as the initial predicted value sample Step 3. for i = 1 to K, do

Step 4. Fit a tree $\hat{f}\times i\ (x)$ with D splits (d + 1 terminal nodes) to the training data

Step 5. Update $\hat{f}(x)$ by adding in a shrunken version of the new tree:

Step 6. $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda\ \hat{f}\times i\ (x)$

Step 7. Update the residuals, $r_i \leftarrow r_i - \lambda\ \hat{f}\times i\ (x)$

Step 8. End

III. TRAINING AND TESTING DATASETS

The dataset is split into training, validation and testing sets. The training set consists of 43,355 data values, and it is split into training and validation in the ratio of 80% to 20%, respectively. The training set data is used to train the model, and finally, the testing set is used to test the performance. The data splitting is shown in Table 3. The experiments are carried out at constant temperature in the chamber at 25 ∘C. Increase or decrease in the rising temperature affects the performance of the battery.

Table 3. Battery dataset splitting.

Dataset Splitting		
Total Training Set = 43,355		Testing Set
Training set (80%) 34,684	Validation set (20 %) 8671	25,416

Dataset Splitting

Figures 6 and 7 show the corresponding battery voltage, battery current, battery capacity and battery state of charge profiles of training and testing data. The data used include the voltage, current and SoC values reported during a Panasonic 18650PF battery test [38]. All obtained data are normalized to reduce the fluctuation in training and also the speed of the training time. The Bayesian optimization algorithm optimizes the model using hyperparameter tuning of the proposed machine learning. Typically, this algorithm requires more time, but it can result in good generalization for difficult, small or noisy datasets. According to adaptive weight minimization, training stops (regularization). The Levenberg–Marquardt backpropagation algorithm takes more memory but less time. Training stops automatically when generalization stops improving, as shown by a rise in the validation samples' mean square error.
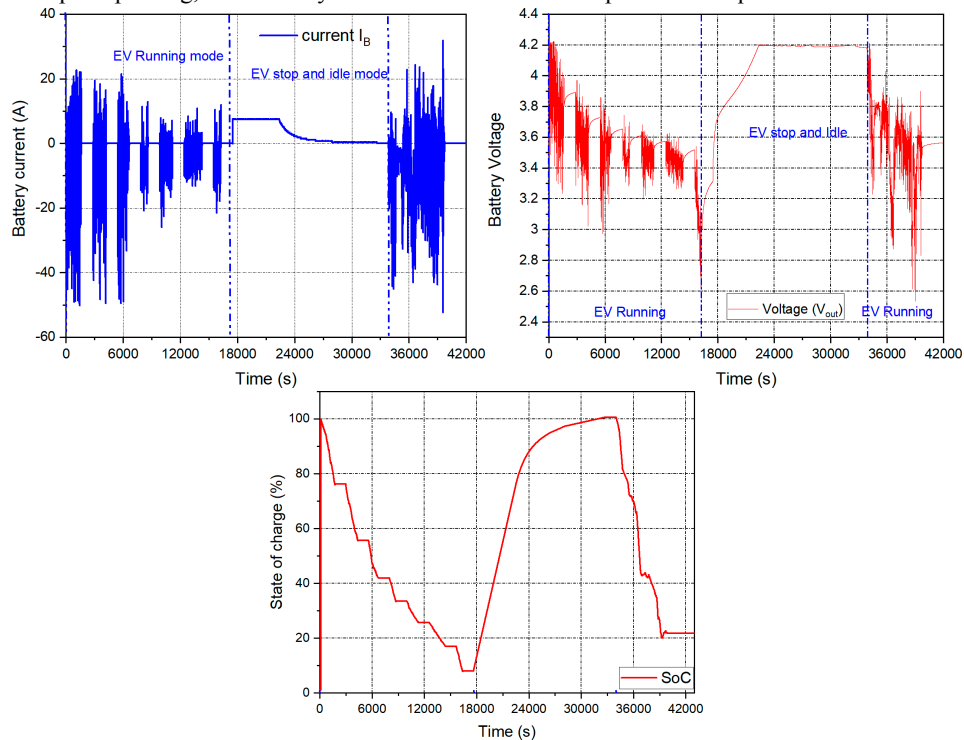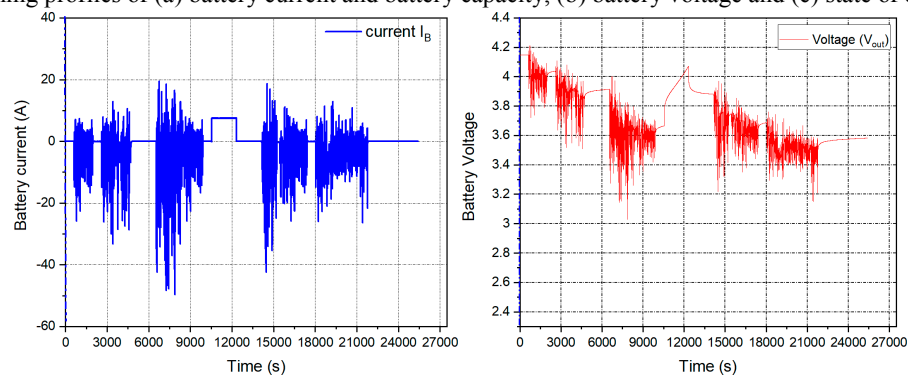

Figure 6. Training profiles of (a) battery current and battery capacity, (b) battery voltage and (c) state of charge.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/568
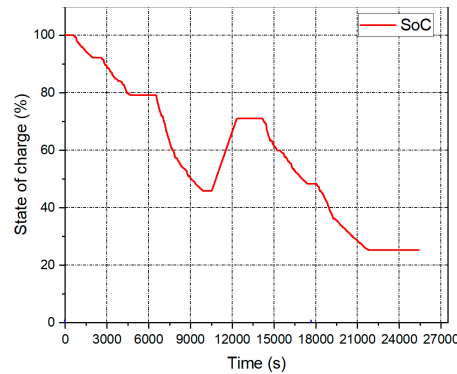
ISSN
2581-9429
IJARSCT

113

Figure 7. Testing dataset of (a) battery current and battery capacity, (b) battery voltage and (c) state of charge (SoC). The Bayesian optimization in the neural network provides the optimal neurons and number of layers best fitted for the obtained dataset by considering its RMSE value. In ensemble boosting and bagging machine learning, the Bayesian optimization algorithm plays a major role in selecting the tree's depth to fine-tune the model. As defined in Section 2, the artificial neural network is one of the main algorithms in machine learning.

IV. PERFORMANCE METRICS

To evaluate the results of our the predicted SoC for the adopted models, we need to compare the predicted SoC with the experiment SoC's actual results. The performance metrics are therefore assessed by these different metrics [60].

4.1. Root Mean Square Error (RMSE)

The root mean square error is simply the square root of the square mean of all the errors. RMSE is a good measure of accuracy, but only applicable to comparing model predictions with data and not between variables.

$$RMSE = \sum_{i=1}^{m} \frac{\left((Y_{predicted})_i - (Y_{measured})_i \right)^2}{N} \tag{8}$$

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \frac{\left((Y_{predicted}) - (Y_{measured})_i \right)^2_i}{(Y_{measured})_i} \tag{9}$$

4.2. R^2 Square

This is a statistical indicator that describes the amount of uncertainty explained by an independent variable.

$$r^2 = \left[\frac{n(\sum zy) - (\sum z)(\sum y)}{\sqrt{n \sum z^2 - (\sum z)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \right]$$

V. RESULTS AND DISCUSSION

The neural network model is a learning-prediction method in battery management systems for an EVs, such as learning the SoC relationship based on the charging and discharging process of data and then using it to predict the real-time SoC relationship under realistic operating conditions. As seen in the Figure 8a, each charging and discharging count time is fed into the ANN (input-size = 4, hidden-size = 10), and then the data function becomes 4. According to the predicted model, the ANN hidden-size function can be modified: the larger the hidden-size, the more accurate the computational effort. After creating an artificial neural network model, the three features flow into the linear layer and are transformed into one feature. This section analyses the efficiency of SoC's ML forecast model.
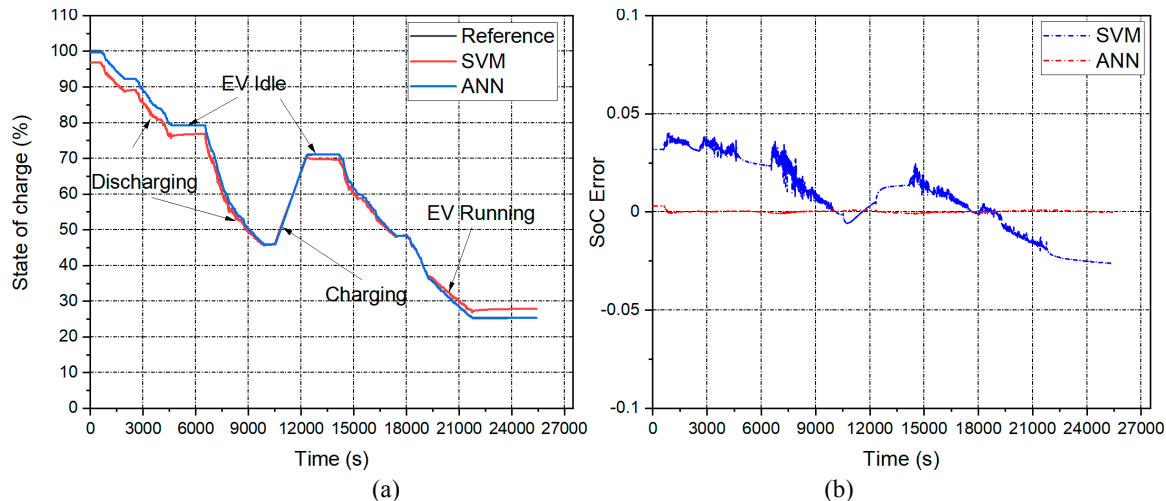
Figure 8. Comparison of SoC estimation methods: (a) support vector machine (SVM) and artificial neural network (ANN) algorithm and (b) SoC error of SVM and ANN.

Figure 8a shows the predicted SoC estimation of the support vector machine and artificial neural network with the expected state of charge. The neural network has a better prediction rate than the support vector machine due to its capability to handle non-linear data. Figure 8b shows the error plot of the neural network and support vector machine for predicted and actual SoC. The SoC measurement methods' overall performance assessment results under four conditions in various modes of electric vehicle operation are shown in Figure 8a.

Figure 9a compares the precited values of SoC of gaussian process regression and linear regression for the actual SoC estimation value. The Gaussian process regression has a better advantage over the linear regression. The Gaussian process can give the most reliable prediction of their uncertainty. However, the GPR will require more training time when compared to the linear regression as it takes the entire training dataset for training. The error analysis of the SoC is shown in Figure 9b. The SoC measurement methods' overall performance assessment results under four conditions in various electric vehicle operation modes are shown in Figure 9a,b. Figure 10a describes the comparative SoC estimation results of the proposed ensemble bagging and ensemble boosting. The ensemble models have the advantages of converting the weak learner to the stronger learner to predict the estimated values. At the ensembled regression model's output, the average value is calculated for the predicted regression values.
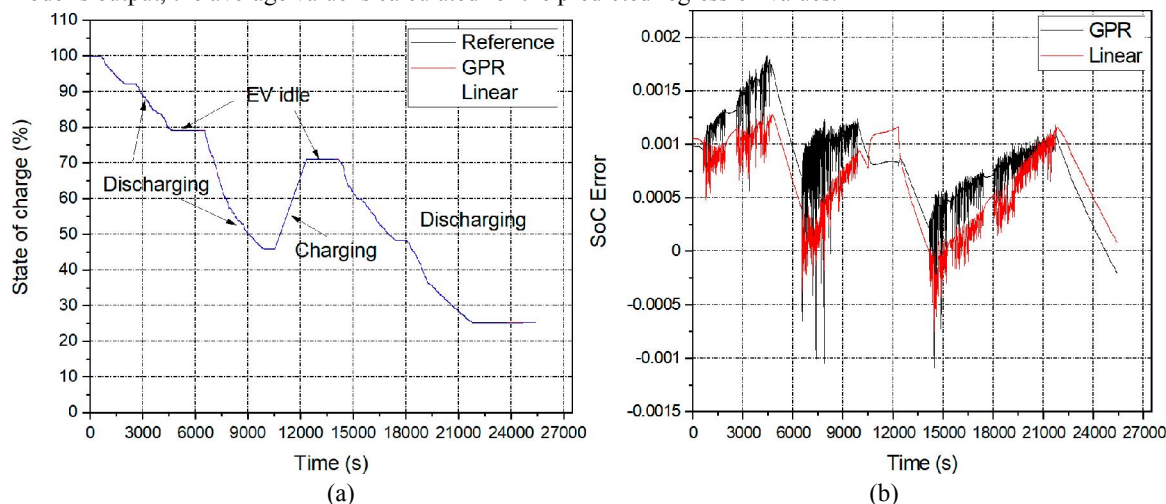

Figure 9. Comparison of SoC estimation methods: (a) Gaussian process regression (GPR) and linear algorithm and (b) SoC error of GPR and linear.

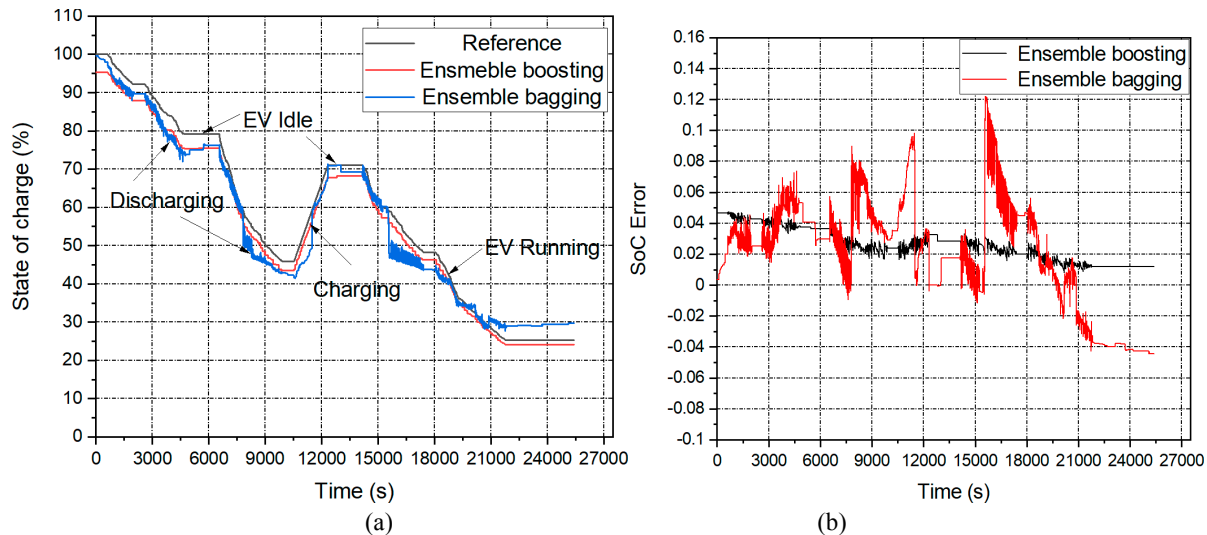(a)   (b)

Figure 10. Comparison of SoC estimation methods: (a) ensemble boosting and ensemble bagging algorithms, (b) SoC error of ensemble boosting and ensemble bagging.

The ensemble boosting prediction is a better SoC estimation when compared with ensemble bagging. The ensemble boosting has the characteristics of working better with multiple input features directly affecting training data. The ensemble bagging is not showing better results due to the problem of overfitting in the training data. The error plot of the proposed method is shown in Figure 10b. Another amazing detail in Figure 10 is that the ensemble bagging and boosting system accomplished low efficiency with great fluctuations and error. This is presumably because the temporal dependence between historical measurements and the SoC is not considered for the ensemble process. Finally, the learned model is obtained by repeating the learning process until the error is within a reasonable range. Blue lines and red lines are the projected SoC values and the true SoC values, and the light blue regions are the confidence level of the estimated SoC values.

Table 4 outlines the performance analysis of the different proposed machine learning algorithms. Table 4 and Figures 8, 9 and 10a,b show that the proposed GPR approach achieves good efficiency with 85% Mean absolute Error(MAE) which outperformed all methods. The suggested GPR-linear approach reduces the MAE by 51% and 50%, respectively. This may be because the GPR kernel is capable of collecting the dynamic time structures of sequential data. In comparison, the GPR-linear approach produces a better prediction outcome than the SVM-ANN, with a 10% reduction in the MAE. The experi- ment's findings demonstrate that the proposed approach will accomplish SoC prediction under varying ambient temperature conditions with several network parameters. The standard SVM-ANN approach and the ensemble trees method suggested the lower MAE, variance (VAF) obtained, as seen in Table 4 and Figure 11. One particular benefit of the proposed approach is that it can provide confidence intervals for the SoC calculations and infer the SoC estimation values' volatility. This is critical for evaluating the volatility of the forecasts and thus offers more insightful performance.

Table 4. Performance analysis of the battery.

| Algorithm	MSE	RMSE	NRMSE	MAE	MAPE	Scatter Index	Variance	R2
SVM	0.01505	0.12266	0.17517	0.00752	0.000052	0.21	81.63	0.999
ANN	0.00054	0.02329	0.03126	0.00027	0.000002	0.040	99.99	0.999
Linear	0.00130	0.03610	0.04829	0.00065	0.000004	0.062	99.95	0.979
GPR	0.00170	0.04118	0.05507	0.00085	0.000006	0.071	99.83	1.000
Ensemble boosting	0.05245	0.22902	0.32122	0.02623	0.000186	0.39	90.32	1.000
Ensemble bagging	0.04231	0.04118	0.28576	0.02115	0.000149	0.35	85.25	0.979
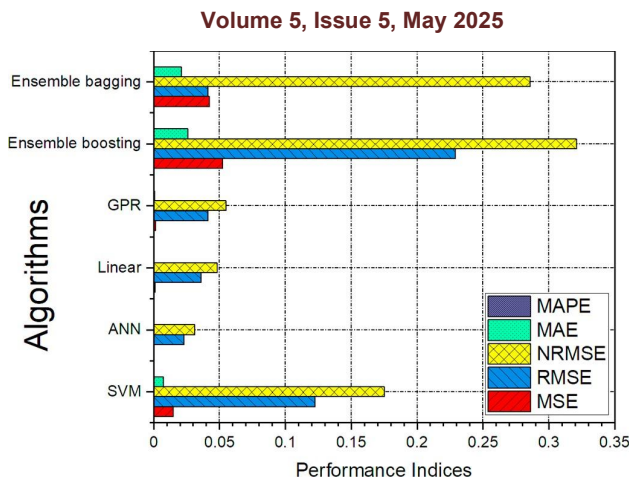
Figure 11. Performance indices of the proposed ML algorithms.

For model training, different numbers of neurons in the ANN model are used. The RMSE of the SoC estimation and the time taken to evaluate the ML model's performance are calculated. The calculation time for ML increases considerably for an increase in the number of neurons in the neural networks. As the number of neurons reaches 10, RMSE and MSE eventually converge to a stable value. It can be shown that, if the statistic is higher than 15, an increase in neurons does not greatly increase the performance of the calculation but loses measuring time. These findings demonstrate that ML does not lead to overfitting because RMSE converges to a small value.

VI. CONCLUSIONS

Prediction of lithium-ion batteries' SoC plays a vital role in the battery management system of the electric vehicle's performance. In this work, the battery SoC is predicted based on six-machine learning algorithms which include artificial neural network (ANN), support vector machine (SVM), linear regression (LR), Gaussian process regression (GPR), ensemble bagging and ensemble boosting algorithms. With the proposed machine learning models, the non-linear mapping of the input features such as voltage and current to the SoC estimation is analyzed. Machine learning algorithms are selected for estimating the battery SoC due to their better handling of non-linear data. Besides, the proposed method can be used for real-time SoC estimations after optimizing the GPR-linear model hyperparameters. With 85% MAE, the proposed ANN and GPR approach achieves strong performance while outperforming other methods. This could be because the GPR kernel can extract sequential data from complex time structures. In contrast, the GPR-linear approach performs better than the SVM-ANN, with a 10% decrease in the MAE. We conclude that the proposed ANN and GPR-based method further encourages improvement in the SoC estimate because of the probability distribution rather than the estimation of the point. The optimized features input into the machine learning model predict the battery state of charge estimation, which will help stakeholders and the researchers to identify their best battery for specific applications. ANN and GPR will help design the optimum battery management system for electric vehicles based on SoC predictions.

REFERENCES

[1]. Ghosh, A. Possibilities and Challenges for the Inclusion of the Electric Vehicle (EV) to Reduce the Carbon Footprint in the Transport Sector: A Review. Energies 2020, 13, 2602. [CrossRef]

[2]. Tang, X.; Gao, F.; Zou, C.; Yao, K.; Hu, W.; Wik, T. Load-responsive model switching estimation for state of charge of lithium-ion batteries. Appl. Energy 2019, 238, 423–434. [CrossRef]

[3]. S´wierczyn´ski, M.; Stroe, D.I.; Laerke, R.; Stan, A.I.; Kjaer, P.C.; Teodorescu, R.; Kaer, S.K. Field Experience from Li-Ion BESS

[4]. Delivering Primary Frequency Regulation in the Danish Energy Market. ECS Trans. 2014, 61, 1–14. [CrossRef]

[5]. Rahimi-Eichi, H.; Ojha, U.; Baronti, F.; Chow, M.Y. Battery management system: An overview of its application in the smart grid and electric vehicles. IEEE Ind. Electron. Mag. 2013, 7, 4–16. [CrossRef]

[6]. Ramanan, P.; Kalidasa Murugavel, K.; Karthick, A.; Sudhakar, K. Performance evaluation of building-integrated photovoltaic systems for residential buildings in southern India. Build. Serv. Eng. Res. Technol. 2019, 41, 492–506. [CrossRef]

[7]. Karthick, A.; Athikesavan, M.M.; Pasupathi, M.K.; Kumar, N.M.; Chopra, S.S.; Ghosh, A. Investigation of inorganic phase change material for a semi-transparent photovoltaic (STPV) module. Energies 2020, 13, 3582. [CrossRef]

[8]. Chandrika, V.S.; Thalib, M.M.; Karthick, A.; Sathyamurthy, R.; Manokar, A.M.; Subramaniam, U.; Stalin, B. Performance assessment of free standing and building integrated grid connected photovoltaic system for southern part of India. Build. Serv. Eng. Res. Technol. 2020. [CrossRef]

[9]. Berecibar, M.; Gandiaga, I.; Villarreal, I.; Omar, N.; Van Mierlo, J.; Van Den Bossche, P. Critical review of state of health estimation methods of Li-ion batteries for real applications. Renew. Sustain. Energy Rev. 2016, 56, 572–587. [CrossRef]

[10]. Cadini, F.; Sbarufatti, C.; Cancelliere, F.; Giglio, M. State-of-life prognosis and diagnosis of lithium-ion batteries by data-driven particle filters. Appl. Energy 2019, 235, 661–672. [CrossRef]

[11]. Bhattacharjee, A.; Mohanty, R.K.; Ghosh, A. Design of an Optimized Thermal Management System for Li-Ion Batteries under Different Discharging Conditions. Energies 2020, 13, 5695. [CrossRef]

[12]. Karthick, A.; Kalidasa Murugavel, K.; Ghosh, A.; Sudhakar, K.; Ramanan, P. Investigation of a binary eutectic mixture of phase change material for building integrated photovoltaic (BIPV) system. Sol. Energy Mater. Sol. Cells 2020, 207. [CrossRef]

[13]. Pagani, M.; Korosec, W.; Chokani, N.; Abhari, R.S. User behaviour and electric vehicle charging infrastructure: An agent-based model assessment. Appl. Energy 2019, 254. [CrossRef]

[14]. Karthick, A.; Kalidasa Murugavel, K.; Suse Raja Prabhakaran, D. Energy analysis of building integrated photovoltaic modules. In Proceedings of the International Conference on Power and Embedded Drive Control, ICPEDC 2017, Chennai, India, 16–18 March 2017; pp. 307–311.

[15]. Ghosh, A. Potential of building integrated and attached/applied photovoltaic (BIPV/BAPV) for adaptive less energy-hungry building's skin: A comprehensive Review. J. Clean. Prod. 2020, 123343. [CrossRef]

[16]. Reddy, P.; Gupta, M.V.N.S.; Nundy, S.; Karthick, A. Status of BIPV and BAPV System for Less Energy-Hungry Building in India—A Review. Appl. Sci. 2020, 10, 2337. [CrossRef]

[17]. Khalid, M.; Shanks, K.; Ghosh, A.; Tahir, A.; Sundaram, S.; Mallick, T.K. Temperature regulation of concentrating photovoltaic window using argon gas and polymer dispersed liquid crystal fi lms. Renew. Energy 2021, 164, 96–108. [CrossRef]

[18]. Mesloub, A.; Ghosh, A. Daylighting performance of light shelf photovoltaics (LSPV) for office buildings in hot desert-like regions.Appl. Sci. 2020, 10, 7959. [CrossRef]

[19]. Mesloub, A.; Ghosh, A.; Touahmia, M. Performance Analysis of Photovoltaic Integrated Shading Devices (PVSDs) and Semi- Transparent Photovoltaic (STPV) Devices Retrofitted to a Prototype O ffi ce Building in a Hot Desert Climate. Sustainability 2020, 12, 10145. [CrossRef]

[20]. Kotia, A.; Borkakoti, S.; Ghosh, S.K. Wear and performance analysis of a 4-stroke diesel engine employing nanolubricants.Particuology 2018, 37, 54–63. [CrossRef]

[21]. Ramanan, P.; Karthick, A. Performance analysis and energy metrics of grid-connected photovoltaic systems. Energy Sustain. Dev.2019, 52, 104–115. [CrossRef]

[22]. Amjad, M.; Ahmad, A.; Rehmani, M.H.; Umer, T. A review of EVs charging: From the perspective of energy optimization, optimization approaches, and charging techniques. Transp. Res. Part D Transp. Environ. 2018, 62, 386–417. [CrossRef]

[23]. Sathyamurthy, R.; Kabeel, A.E.; Chamkha, A.; Karthick, A.; Muthu Manokar, A.; Sumithra, M.G. Experimental investigation on cooling the photovoltaic panel using hybrid nanofluids. Appl. Nanosci. 2020. [CrossRef]

[24]. Ghosh, A.; Norton, B.; Duffy, A. First outdoor characterisation of a PV powered suspended particle device switchable glazing.Sol. Energy Mater. Sol. Cells 2016, 157, 1–9. [CrossRef]

[25]. Ghosh, A.; Norton, B. Optimization of PV powered SPD switchable glazing to minimise probability of loss of power supply.Renew. Energy 2019, 131, 993–1001. [CrossRef]

[26]. Liu, K.; Shang, Y.; Ouyang, Q.; Widanage, W.D. A Data-driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery. IEEE Trans. Ind. Electron. 2020, 1. [CrossRef]

[27]. Bonfitto, A.; Ezemobi, E.; Amati, N.; Feraco, S.; Tonoli, A.; Hegde, S. State of health estimation of lithium batteries for automotive applications with artificial neural networks. In Proceedings of the 2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Turin, Italy, 2–4 July 2019; pp. 1–5. [CrossRef]

[28]. Dang, X.; Yan, L.; Xu, K.; Wu, X.; Jiang, H.; Sun, H. Open-Circuit Voltage-Based State of Charge Estimation of Lithium-ion Battery Using Dual Neural Network Fusion Battery Model. Electrochim. Acta 2016, 188, 356–366. [CrossRef]

[29]. Chang, Y.; Fang, H.; Zhang, Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery.Appl. Energy 2017, 206, 1564–1578. [CrossRef]

[30]. Nuhic, A.; Terzimehic, T.; Soczka-Guth, T.; Buchholz, M.; Dietmayer, K. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. J. Power Sources 2013, 239, 680–688. [CrossRef]

[31]. Dai, H.; Zhao, G.; Lin, M.; Wu, J.; Zheng, G. A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and markov chain. IEEE Trans. Ind. Electron. 2019, 66, 7706–7716. [CrossRef]

[32]. Feng, X.; Weng, C.; He, X.; Han, X.; Lu, L.; Ren, D.; Ouyang, M. Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine. IEEE Trans. Veh. Technol. 2019, 68, 8583–8592. [CrossRef]

[33]. Liu, K.; Li, Y.; Hu, X.; Lucu, M.; Widanage, W.D. Gaussian Process Regression with Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-Ion Batteries. IEEE Trans. Ind. Inform. 2020, 16, 3767–3777. [CrossRef]

[34]. Liu, K.; Hu, X.; Wei, Z.; Li, Y.; Jiang, Y. Modified Gaussian Process Regression Models for Cyclic Capacity Prediction of Lithium-Ion Batteries. IEEE Trans. Transp. Electrif. 2019, 5, 1225–1236. [CrossRef]

[35]. Richardson, R.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for forecasting battery state of health. arXiv2017, 357, 209–219. [CrossRef]

[36]. Malkhandi, S. Fuzzy logic-based learning system and estimation of state-of-charge of lead-acid battery. Eng. Appl. Artif. Intell.2006, 19, 479–485. [CrossRef]

[37]. Burgos, C.; Sáez, D.; Orchard, M.E.; Cárdenas, R. Fuzzy modelling for the state-of-charge estimation of lead-acid batteries.J. Power Sources 2015, 274, 355–366. [CrossRef]

[38]. Lee, D.T.; Shiah, S.J.; Lee, C.M.; Wang, Y.C. State-of-charge estimation for electric scooters by using learning mechanisms.IEEE Trans. Veh. Technol. 2007, 56, 544–556. [CrossRef]

[39]. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Emadi, A. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. J. Power Sources 2018, 400, 242–255. [CrossRef]

[40]. Ting, T.O.; Man, K.L.; Lim, E.G.; Leach, M. Tuning of Kalman filter parameters via genetic algorithm for state-of-charge estimation in battery management system. Sci. World J. 2014, 2014. [CrossRef]

[41]. Li, I.H.; Wang, W.Y.; Su, S.F.; Lee, Y.S. A merged fuzzy neural network and its applications in battery state-of-charge estimation.IEEE Trans. Energy Convers. 2007, 22, 697–708. [CrossRef]

[42]. Chen, L.; Wang, Z.; Lu, Z.; Li, J.; Ji, B.; Wei, H.; Pan, H. A novel state-of-charge estimation method of lithium-ion batteries combining the grey model and genetic algorithms. IEEE Trans. Power Electron. 2018, 33, 8797–8807. [CrossRef]

[43]. Bonfitto, A.; Feraco, S.; Tonoli, A.; Amati, N.; Monti, F. Estimation accuracy and computational cost analysis of artificial neural networks for state of charge estimation in lithium batteries. Batteries 2019, 5, 47. [CrossRef]

[44]. Xu, Z.; Wang, J.; Fan, Q.; Lund, P.D.; Hong, J. Improving the state of charge estimation of reused lithium-ion batteries by abating hysteresis using machine learning technique. J. Energy Storage 2020, 32. [CrossRef]

[45]. Liu, Y.; Guo, B.; Zou, X.; Li, Y.; Shi, S. Machine learning assisted materials design and discovery for rechargeable batteries. Energy Storage Mater. 2020, 31, 434–450. [CrossRef]

[46]. Kim, S.; Lim, H. Reinforcement learning based energy management algorithm for smart energy buildings. Energies 2018, 11, 2010. [CrossRef]

[47]. Attia, M.E.H.; Karthick, A.; Manokar, A.M.; Driss, Z.; Kabeel, A.E.; Sathyamurthy, R.; Sharifpur, M. Sustainable potable water production from conventional solar still during the winter season at Algerian dry areas: Energy and exergy analysis. J. Therm. Anal. Calorim. 2020. [CrossRef]

[48]. Dhanalakshmi, C.S.; Madhu, P.; Karthick, A.; Mathew, M.; Vignesh Kumar, R. A comprehensive MCDM-based approach using TOPSIS and EDAS as an auxiliary tool for pyrolysis material selection and its application. Biomass Convers. Biorefinery 2020. [CrossRef]

[49]. Karthick, A.; Ramanan, P.; Ghosh, A.; Stalin, B.; Vignesh Kumar, R.; Baranilingesan, I. Performance enhancement of copper indium diselenide photovoltaic module using inorganic phase change material. Asia-Pac. J. Chem. Eng. 2020, 15. [CrossRef]

[50]. Shepero, M.; Munkhammar, J.; Widén, J.; Bishop, J.D.K.; Boström, T. Modeling of photovoltaic power generation and electric vehicles charging on city-scale: A review. Renew. Sustain. Energy Rev. 2018, 89, 61–71. [CrossRef]

[51]. Hoarau, Q.; Perez, Y. Interactions between electric mobility and photovoltaic generation: A review. Renew. Sustain. Energy Rev. 2018, 94, 510–522. [CrossRef]

[52]. Karthick, A.; Kalidasa Murugavel, K.; Sudalaiyandi, K.; Muthu Manokar, A. Building integrated photovoltaic modules and the integration of phase change materials for equatorial applications. Build. Serv. Eng. Res. Technol. 2020, 41, 634–652. [CrossRef]

[53]. Sudalaiyandi, K.; Alagar, K.; VJ, M.P.; Madhu, P. Performance and emission characteristics of diesel engine fueled with ternary blends of linseed and rubber seed oil biodiesel. Fuel 2021, 285, 119255. [CrossRef]

[54]. Li, Y.; Sheng, H.; Cheng, Y.; Stroe, D.I.; Teodorescu, R. State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis. Appl. Energy 2020, 277. [CrossRef]

[55]. Mawonou, K.S.R.; Eddahech, A.; Dumur, D.; Beauvois, D.; Godoy, E. State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. J. Power Sources 2020. [CrossRef]

[56]. Yang, D.; Zhang, X.; Pan, R.; Wang, Y.; Chen, Z. A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. J. Power Sources 2018, 384, 387–395. [CrossRef]

[57]. Lu, C.; Tao, L.; Fan, H. Li-ion battery capacity estimation: A geometrical approach. J. Power Sources 2014, 261, 141–147. [CrossRef]

[58]. Shu, X.; Li, G.; Zhang, Y.; Shen, J.; Chen, Z.; Liu, Y. Online diagnosis of state of health for lithium-ion batteries based on short-term charging profiles. J. Power Sources 2020, 471. [CrossRef]

[59]. Li, W.; Jiao, Z.; Du, L.; Fan, W.; Zhu, Y. An indirect RUL prognosis for lithium-ion battery under vibration stress using Elman neural network. Int. J. Hydrogen Energy 2019, 44, 12270–12276. [CrossRef]

[60]. Stroe, D.I.; Schaltz, E. Lithium-Ion Battery State-of-Health Estimation Using the Incremental Capacity Analysis Technique. IEEE Trans. Ind. Appl. 2020, 56, 678–685. [CrossRef]

[61]. Li, Y.; Zou, C.; Berecibar, M.; Nanini-Maury, E.; Chan, J.C.W.; van den Bossche, P.; Van Mierlo, J.; Omar, N. Random forest regression for online capacity estimation of lithium-ion batteries. Appl. Energy 2018, 232, 197–210. [CrossRef]