20th IMEKO TC10 Conference
"Measurement for Diagnostics, Optimization and Control to Support Competitiveness and Innovation"
Lisbon, Portugal, September 28–29, 2026

State of Charge Estimation of Energy Storage Systems using a Flexible Reward based Reinforcement Learning Approach

Sadia Ali[1], Valentina Bianchi[1], Gabriele Patrizi[2], Lorenzo Ciani[2], Ilaria De Munari[1]

[1]*Università degli studi di Parma, Parma, Italy, sadia.ali@unipr.it, valentina.bianchi@unipr.it, ilaria.demunari@unipr.it*
[2]*Università degli studi di Firenze, Firenze, Italy, gabriele.patrizi@unifi.it, lorenzo.ciani@unifi.it*

***Abstract* – A flexible reward-based reinforcement learning (RL) approach for the state of charge (SoC) estimation of energy storage systems (ESSs) i.e. a Lithium-ion battery (LiB) and a hybrid supercapacitor (HSC), is presented in this research. The data acquisition for the Panasonic 18650 LiB is carried out through constant current – constant voltage (CC-CV) procedure whereas, for the cylindrical HSC electrochemical impedance spectroscopy (EIS) is used. A three-stage approach is presented i.e. a reinforcement learning (RL) framework followed by a least square boosting (LSB) stage for the correction of residual errors from the RL, and a three-point linear interpolation final processing. RL incorporates a twin-delayed deep deterministic (TD3) neural network-based policy gradient agent along with a reward function specifically designed to take into account a flexible use in possible different application contexts. The RMSE, along with other performance metrics, is reported and analysed for each discharge cycle in the test datasets. The overall minimum average RMSE per cycle for the LiB and HSC are reported to be 0.87%, 0.33% whereas the overall average RMSEs are 1.04% and 0.44%, respectively. The ESSs results are progressive in comparison with the existing literature and validate the efficacy of the multi-term flexible reward function approach.**

Keywords – Electrochemical impedance spectroscopy; Reinforcement learning; Lithium-ion battery; Ensemble learning; State-of-charge; Hybrid Supercapacitor; Sustainable Cities and Communities.

I. INTRODUCTION

The shift towards sustainable cities and communities requires minimum usage of fuel-based vehicles and an encouraging leap towards electric vehicles (EVs) [1]. Moreover, a rapid decline in the availability of conventional energy sources, makes this shift obligatory. The major component within the EV is the energy storage system (ESS), which requires an efficient management system to deal with parameter estimations and cell balancing [2]. One of the most inevitable factors in the transformation from fuel-based vehicles to EVs, is the accurate estimation of the state of charge (SoC) of the ESS, given that an inaccurate estimation can cause unexpected and potentially dangerous real-time situations [3].

Among the various ESSs, the lithium-ion battery (LiB) is the most widely used and experimented on device [4]. For the SoC estimation of a LiB, an improved algorithm based on Coulomb Counting and uncertainty evaluation, considering a period of ten years, is presented in [5]. This mathematical model-based approach results in a percentage error of 0.3%. In [6], another SoC estimation technique based on an improved bidirectional gated recurrent unit (BGRU) and an unscented Kalman filter (UKF) is proposed for the SoC estimation of a single ESS, i.e. LiB. The approach uses voltage, current and temperature as input features under UDDS and US06 driving cycles. The method is validated for different temperatures and results in an RMSE less than 1.12%.

Another ESS, the supercapacitor (SC) stands out due to its high energy density, high power density, fast charging, rapid discharge and long cycle life [7]. A sub-type of SC, the hybrid supercapacitor (HSC) sits between a battery and a traditional SC, naturally uniting the qualities of both these devices [8]. Due to its advantages, the HSC is now more aggressively researched on for the integration with EVs. For the SoC estimation of an HSC, a UKF based on a fractional order model (FO-AUKF) is presented in [9], addressing a joint estimation phenomenon for SoC and State of Power (SoP). The fractional order-based model is combined with a competitive learning-based particle swarm optimization (CLPSO) algorithm. The CLPSO algorithm is deployed for parameter identification whereas the FO-AUKF is deployed for the state estimation. The approach results in a SoC estimation with an RMSE of less than 2% at temperatures ranging from -10 °C to 45 °C. Another article [10] proposes a switched system approach designed specifically for the RC model of the supercapacitor (SC), to be used mainly while working with cell imbalances within an ESS. This approach results in a tracking error of less than 1%. An online scheme for an accurate estimation of SoC of an SC is presented in [11]. This approach considers the leakage effect to ensure proper and efficient management of the SC. The technique results in a minimum absolute average percentage error of 0.35%, with a 20% uncertainty introduced in the equivalent series resistance of the SC. A gaussian process regression (GPR) based fast approach incorporating the data collected for a HSC, through electrochemical impedance spectroscopy (EIS) is presented in [12]. The approach offers an optimal balance between accuracy and timing by incorporating two separate frequency points for SoC within [0,30] and [30,100] respectively. The technique results in an average root mean square error (RMSE) of 0.94%.

The RL is an agent-based technique that, based on a reward function, takes consecutive steps, interacting with the environment, to maximize the designed reward [13]. The custom reward function is formulated keeping in view the ultimate objectives derived from the problem statement. In [14], a hybrid SoC estimation framework for UAV lithium-ion batteries by combining Adaptive Extended Kalman Filtering with Deep Reinforcement Learning is proposed. In their approach, a PPO-based policy adaptively tunes the process and measurement covariance matrices of the filter, improving estimation robustness under time-varying flight conditions. However, the method is mainly validated on a specific UAV battery system and real flight profile, which may limit its generalization to other energy storage devices or operating scenarios. Since the PPO policy is trained for a particular battery model, mission profile, and covariance adaptation range, its performance may degrade when applied to different batteries, load patterns, temperatures, or aging conditions without retraining or recalibration. In [15], a deep RL based approach for a nickel-metal hybrid battery is presented, to estimate the SoC of the battery. The deep deterministic policy gradient-based approach utilizes a basic error-based reward function and results in an estimation accuracy of 98.8%. The initial prediction often contains a residual error mainly present due to structural complication and algorithmic limitations [16]. For this reason, in [17] the residual error is lowered deploying a second phase, following the RL, exploiting a least square boosting (LSB) algorithm. LSB utilizes the residual errors from the first phase of the training to correct the difference that was not captured during the initial phase i.e. RL. Moreover, following the LSB, a three-point linear interpolation technique is deployed, which directs whether to keep the predicted value or replace it with a computed value by setting a reasonable threshold. However, in this case the reward function was quite simple and targeted to the specific dataset used in the training, based on EIS measurements on an HSC device. The use of the same model on different devices or application contexts would require different approaches in the design of the reward function, limiting the generalizability in other contexts.

Hence, the possibility to explore RL for SoC assessment, designing a flexible approach that can be easily adapted to different ESS and applications field is particularly interesting and it has not been studied yet. In this paper, a flexible data driven approach based on RL has been proposed and implemented for two different ESSs, a LiB and an HSC. The aim is to design a custom reward function that, by means of proper parameters, could be adapted to different discharge profiles and different ESS. The whole approach also includes two additional stages, a LSB and a three-point linear interpolation to further improve the accuracy in the estimation of the SoC and, in particular, the RMSE metric.

The paper is organized as follows: in Section II the methodology for acquiring the data used in the training and test processes is detailed; in Section III the proposed framework is described; in Section IV the results are presented and discussed, and in Section V the conclusions are drawn.

II. DATA ACQUISITION

A. LITHIUM-ION BATTERY (LiB):

The first dataset is based on a Panasonic NCR18650PF, a 2.75 Ah cylindrical Li-ion battery with nickel-manganese-cobalt oxide chemistry, nominal voltage of 3.6 V, operating voltage limits of 2.5-4.2 V, and a maximum discharge current of 10 A [18]. Data acquisition is conducted using a standard constant-current/constant-voltage (CC-CV) procedure, in accordance with the battery datasheet specifications, to ensure that each discharge cycle starts from a fully charged cell and hence repeatable initial condition. The charging phase is followed by a 30 min rest period before each subsequent discharge phase in order to reduce polarization effects and make consecutive cycles independent. Voltage, current, and temperature are acquired through a laboratory workbench consisting of an ITECH IT-M3412 bidirectional programmable power supply, operated as both charger and electronic load, and three Agilent 34401A digital multimeters connected to a PC through a USB/GPIB interface. The first DMM measures the cell voltage in a 10 V range with 100 μV resolution, the second estimates the discharge current from the voltage drop across a 10 mΩ shunt resistor, and the third measures the cell surface temperature through a 10 kΩ NTC thermistor mounted longitudinally on the battery holder. The acquisition system is controlled through MATLAB and obtains one measurement every 100 ms.

The numerical values of current, voltage, and temperature are collected over multiple discharge cycles starting at ambient temperature, without temperature control simulating realistic conditions. For the purpose of this paper, five repeated profiles including currents in the range of 2–3 A, with a 0.1 A increment, have been selected. The reference SoC values are obtained through Coulomb counting under controlled laboratory conditions.

For model development, the training and testing data are randomized by creating five randomized datasets (RDs), each of which is divided into 80% training and 20% test sub-datasets. For each RD, the proposed framework is trained and evaluated independently, resulting in five training instances and allowing the robustness of the estimation method to be assessed across different randomized data partitions.

B. HYBRID SUPERCAPACITOR (HSC):

The second dataset used in this work includes EIS measurements acquired from a battery-like cylindrical 4.2 V, 4000 F rechargeable HSC, operated within the manufacturer-recommended non-stressful current limits in order to avoid accelerated aging and preserve repeatability across cycles [19]. The experimental campaign is designed to collect impedance information at different SoC levels by repeatedly charging the HSC to 100% SoC using a CC-CV strategy, followed by a 30 min resting period to allow electrochemical stabilization and reduce polarization effects. The HSC is then discharged using a stepwise constant-current profile, where each discharge step corresponds to a 2% variation in SoC. Each discharge step is followed by an additional short rest period before performing the EIS measurement, ensuring that the impedance response is acquired under stable conditions. These EIS measurements are repeated over multiple discharge cycles under controlled environmental conditions, with temperature and humidity kept constant to ensure uniformity of the data and reduce the influence of external factors on the impedance spectra.

Potentiostatic EIS is performed at five different frequencies (i.e., 50 mHz, 100 mHz, 500 mHz, 1 Hz, 5 Hz) for multiple discharge cycles and the impedance-related features, i.e. modulus and phase, are used for training the proposed network. For the training and testing procedure, the acquired EIS data are randomized to reduce bias associated with the order of the experimental cycles. Three different RDs are generated, and each RD is divided into training and test subsets using a 70%–30% split, respectively. For each RD, the proposed network is trained and evaluated using the EIS features extracted at the selected frequency points, resulting in a total of 15 training instances.

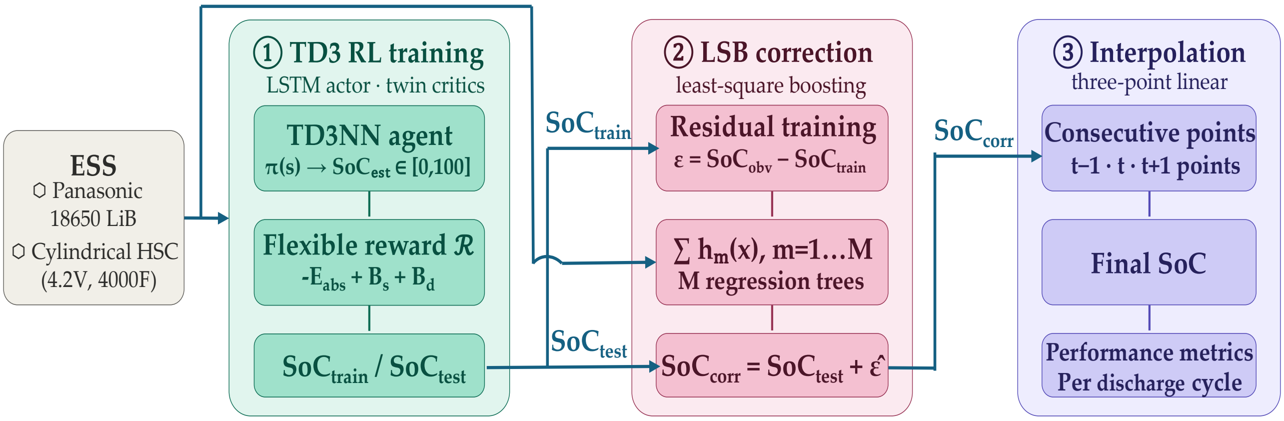
III. PROPOSED FRAMEWORK

In this research work, a three-stage framework for the accurate SoC estimation of ESSs is presented. In the first stage, RL is utilized as the primary estimation technique and has been trained on both datasets (LiB/HSC) separately. The RL framework is based on a twin-delayed deep deterministic neural network (TD3NN) agent with 50 hidden units, an LSTM based architecture. TD3 offers several advantages over conventional methods. It improves training stability by using two critic networks to reduce overestimation of the action-value function. Also, the delayed policy update prevents the actor network from being updated too frequently based on inaccurate value estimates and leads to smoother and more reliable learning. Finally, a specific target policy smoothing helps to avoid exploiting narrow errors in the value function and improves generalization.

MATLAB RL application has been used for the training, with an actor optimizer learning rate of 0.001 sec and a critic optimizer learning rate of 0.01 sec. The second phase incorporating an LSB, learns the residual error patterns by deploying an ensemble of regression trees and consequently adds a predicted residual correction value to improve accuracy [12]. The LSB's role is to correct the residual leftover error from the RL's initial estimate. It starts with computing the residual error as

$$\varepsilon = SoC_{obv} - SoC_{train} \tag{1}$$

where SoC_{obv} is the observed value, and SoC_{train} is the value estimated through TD3NN agent on the training data. This residual is then modeled by an ensemble of M regression trees, where each tree learns to predict a portion of the remaining error rather than the SoC itself. Aggregating the outputs of these trees yields the estimated residual $\hat{\varepsilon}$, which is then added to the initial SoC estimate (i.e., estimated through the TD3NN agent on the test data). The LSB utilizes different numbers of maximum tree splits and learning cycles. The final phase of three-point linear interpolation smoothens out the cycle boundary discontinuities and attempts to balance out any remaining anomalies. This final phase is deployed only during inference operations. The detailed workflow of the proposed approach is presented in Fig. 1 below.



E_{abs} : Absolute SoC error · $\pi(s)$: Policy function evaluated at state s · B_s : smoothness bonus · B_d : descending bonus · SoC_{train} : estimated SoC on training data through RL · SoC_{test} : estimated SoC on test data through RL · ε : residual error · $\hat{\varepsilon}$: estimated residual error · $h_m(x)$: m_{th} regression tree · SoC_{corr} : corrected SoC through LSB

Fig. 1. Detailed workflow of the proposed approach

It is worth noting that the RL training is conducted with an innovative flexible reward function that penalizes estimation error all the while promoting temporal smoothness and a predictable SoC discharge pattern. The agent attempts to maximize the tailored flexible reward function mentioned below in eqs. (1–6).

$$Reward = -\frac{|SoC_{obv} - SoC_{est}|}{E_{\mathrm{m}}} + B_s + B_d \tag{2}$$

$$B_s, \quad \omega_L \leq |SoC_{est} - SoC_{pr}| \leq \omega_H \tag{3}$$

$$\omega_L = \Delta_{min} - 1, \ \omega_H = \Delta_{max} + 1 \tag{4}$$

$$\Delta_{min} = \min[abs(SoC_{obv}(t) - SoC_{obv}\,(t-1))] \tag{5}$$

$$\Delta_{max} = \max[abs(SoC_{obv}(t) - SoC_{obv}(t-1))] \tag{6}$$

$$B_d, \ if \ SoC_{est} < SoC_{pr} \tag{7}$$

Where SoC_{obv} is the observed value, SoC_{est} is, the estimated value at t and SoC_{pr} is the estimated value at t-1, Δ_{min} and Δ_{max} are the absolute minimum and maximum step change between the consecutive points of the observed SoC at time t and t-1. The B_s and B_d are scalar values that provide additional reward spikes when the difference between the current and previous value is within the tolerance window (ω_L, ω_H) and current SoC value is smaller than the previous value, respectively. The tolerance window defines the acceptable range of variation between two consecutive SoC values, computed using the $[\ \Delta_{min}, \Delta_{max}]$. Hence the reward function encourages the estimated SoC value to lie within the tolerance window which mimics the natural rate of change of the measured SoC, therefore instilling temporal consistency and discouraging unrealistic fluctuations. This ensures a smoother and physically plausible SoC estimate. B_s and B_d are selected through hit and trial tuning, by analyzing the effect of the candidate values on the overall training process and reward maximization. The final values of B_s and B_d are selected to be 0.5 and 0.2, respectively. The reward function is flexible, since the tolerance window is dependent on the particular ESS, and discharge cycle. The computation is carried out directly from the observed SoC profile instead of using any predefined or strict fixed thresholds. Since different ESSs, under different operating conditions and load profiles exhibit different rates of SoC variation, by adapting the proposed reward function strategy, the proposed framework can be applied to multiple datasets and application scenarios without any extensive mathematical level restructuring.

IV. RESULTS AND DISCUSSIONS

The proposed framework is deployed for the accurate SoC estimation of three RDs at five different frequencies for the HSC and five RDs for the LiB. The training and tests are performed on MATLAB R2023b on an 8th generation, Intel core i9 with 16 GB RAM system. The performance metrics such as RMSE and mean absolute error (MAE), mean squared error (MSE) and Coefficient of Determination (R2) are computed and analyzed. For the LiB dataset, the values of the four considered metrics are computed for each test cycle in each RD. Then, the values obtained have been averaged to obtain an overall result. For the HSC dataset, the same procedure described above has been repeated, considering all five frequencies separately. The results are summarized in Table I below.

The obtained minimum average RMSE per cycle was 0.87% for the LiB in the RD#1 and 0.33% for the HSC in the RD#2 @ 0.05Hz.

Table I. Average per cycle performance metrics for ESSs

ESS	Frequency (Hz)	RMSE (%)	MAE (%)	MSE (%)	R^2
LiB	-	1.04	1.29	0.96	0.9980
HSC	0.05	0.44	0.25	0.30	0.9997
HSC	0.1	0.55	0.33	0.39	0.9996
HSC	0.5	0.83	0.72	0.61	0.9992
HSC	1	1.00	1.03	0.75	0.9988
HSC	5	0.76	0.60	0.59	0.9993

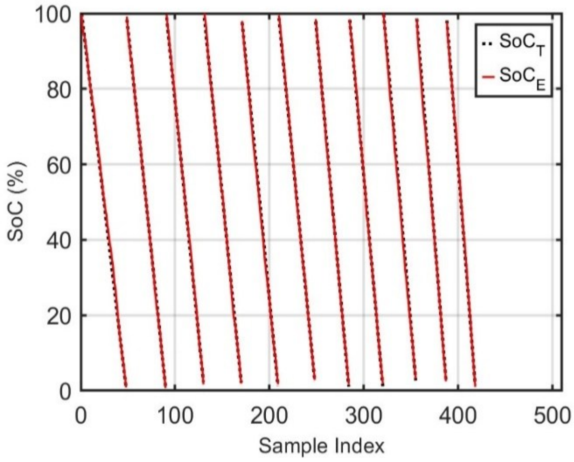


Fig. 2. Example of the estimated SoC vs true SoC (LiB, RD # 1)

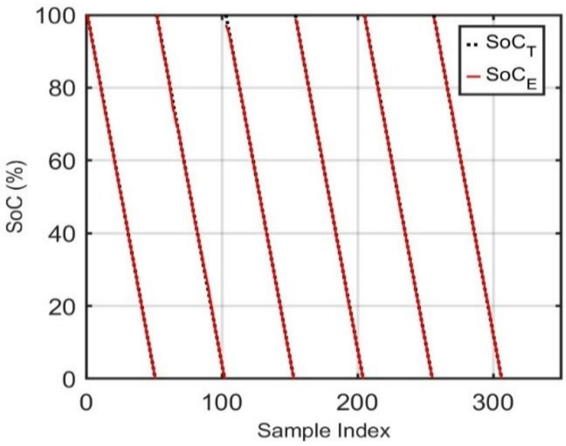


Fig. 3. Example of the estimated SoC vs true SoC (HSC, RD # 2 @ 0.05 Hz)

Table II. Comparison – Existing Literature

ESS	Proposed Technique	RMSE (%)
LiB [6]	BGRU + UKF	0.61
HSC [9]	UKF + particle swarm optimization	0.60
LiB [14]	2RC RL-AEKF	0.44
HSC [17]	RL + LSB + interpolation	0.42
LiB + HSC (proposed)	**Flexible RL + LSB + interpolation**	**0.33**

The results show that the proposed framework achieves accurate SoC estimation for both considered ESS technologies. For the LiB case, the RMSE is 1.04%, with an R2 of 0.9980, indicating that the estimated SoC closely follows the reference trend. However, the MAE of 1.29% suggests that the LiB estimation presents slightly larger average deviations than the HSC cases. For the HSC, the best performance is obtained at 0.05 Hz, where the RMSE and MAE reach 0.44% and 0.25%, respectively, with the highest R2 value of 0.9997. This indicates that low-frequency EIS features provide the most informative impedance response for SoC estimation. At 0.1 Hz, the performance remains very strong, with an RMSE of 0.55% and R2 of 0.9996. As the frequency increases to 0.5 Hz and 1 Hz, the estimation accuracy decreases, reaching the highest RMSE of 1.00% at 1 Hz. This suggests that higher-frequency impedance features may contain less direct information about the SoC-dependent electrochemical dynamics. At 5 Hz, the performance partially improves again, with an RMSE of 0.76%, although it remains less accurate than the lower-frequency cases. Overall, the HSC results are highly frequency-dependent, with the lowest frequencies providing the best estimation accuracy. The consistently high R2 values, always above 0.9988, confirm that the proposed approach can reliably reproduce the SoC trend across all tested cases.

To further show the results, a plot between the true SoC (SoC_T) and the estimated SoC (SoC_E) for the two minimum cases reported are presented above in fig. 2 (i.e. LiB) and fig. 3 (i.e. HSC). Both the plots show promising estimation capabilities of the proposed flexible approach. The estimated SoC trajectory remains visually coherent with the observed SoC with only marginal deviations at the initial transient phases of some of the individual discharge cycles.

A comparison with existing literature is conducted to estimate the efficacy and efficiency of the proposed approach. To ensure a consistent comparison with the studies considered, Table II reports the minimum values achieved by each method, as this performance metric is available across all benchmarks. Our proposed framework provides the minimum RMSE in comparison to the cited approaches. Although a more rigorous comparison would require evaluating all cited methods on a common dataset under identical experimental conditions, the preliminary results highlight the promising potential of the proposed approach. It is worth mentioning that the comparison with [17] is particularly interesting. Indeed, the cited approach utilizes a similar strategy and the same dataset, differing primarily in the reward function. It deploys an absolute negative reward, and the results reveal an improvement of 21% in the RMSE, in favor of our proposed flexible reward approach.

V. CONCLUSIONS

This work presents a flexible three stage RL-based SoC estimation framework for ESSs, incorporating a tailored reward function. The presented framework is applied to one LiB and one HSC with different discharge cycles to prove its adaptable nature. The data extraction is performed through CC-CV and EIS for LiB and HSC, respectively. The three stages include an RL prediction stage based on TD3NN agent, an LSB residual correction stage and an ultimate linear three-point interpolation stage. The RL training is performed with a tailored reward function consisting of additional bonus terms for a descending SoC pattern and to incorporate smoothness. The smoothness bonus is flexible in nature and dependent on the minimum and maximum difference between consecutive values of the observed SoC. Five and three RDs were conducted for the LiB and the HSC, respectively, obtaining a minimum per cycle average RMSE of 0.87% and 0.33%, for LiB and HSC, respectively. The values of these performance metrics when compared with existing literature also prove that the proposed framework provides satisfactory results. The results validate the efficacy of the proposed flexible reward approach and provide a headstart towards replicating this technique for other ESSs as well.

REFERENCES

[1] **R. Kumar, C. Bharatiraja, K. Udhayakumar, S. Devakirubakaran, S. Sathiya, and Lucian Mihet-Popa**, "Advances in Batteries, Battery Modeling, Battery Management System, Battery Thermal Management, SOC, SOH, and Charge/ Discharge Characteristics in EV Applications," *IEEE Access*, vol. 11, pp. 105761–105809, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3318121.

[2] **Popescu, A.**: *Electrical machines and drives*, Politehnium, Iasi, 2011, p. 234.

[3] **Curtley, M, Prince, O.A.**: A synchronous detector with improved parameters, *Proceeding of the International Symposium on Circuits and Systems*, Barcelona, July 15-17, 2009, pp. 255-260.

[4] **W. Chen, J. Liang, Z. Yang, and G. Li**, "A Review of Lithium-Ion Battery for Electric Vehicle Applications and Beyond," *Energy Procedia*, vol. 158, pp. 4363–4368, Feb. 2019, doi: https://doi.org/10.1016/j.egypro.2019.01.783.

[5] **F. Mohammadi**, "Lithium-ion battery State-of-Charge estimation based on an improved Coulomb-Counting algorithm and uncertainty evaluation," *Journal of Energy Storage*, vol. 48, p. 104061, Apr. 2022, doi: https://doi.org/10.1016/j.est.2022.104061.

[6] **Z. Cui, L. Kang, L. Li, L. Wang, and K. Wang**, "A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF," *Energy*, vol. 259, p. 124933, Nov. 2022, doi: https://doi.org/10.1016/j.energy.2022.124933.

[7] **J. Sun, B. Luo, and H. Li**, "A Review on the Conventional Capacitors, Supercapacitors, and Emerging Hybrid Ion Capacitors: Past, Present, and Future," *Advanced Energy and Sustainability Research*, vol. 3, no. 6, p. 2100191, Apr. 2022, doi: https://doi.org/10.1002/aesr.202100191.

[8] **D. P. Dubal, O. Ayyad, V. Ruiz, and P. Gómez-Romero**, "Hybrid energy storage: the merging of battery and supercapacitor chemistries," *Chemical Society Reviews*, vol. 44, no. 7, pp. 1777–1790, 2015, doi: https://doi.org/10.1039/c4cs00266k.

[9] **J. Zhang, B. Xiao, G. Niu, X. Xie, and S. Wu**, "Joint estimation of state-of-charge and state-of-power for hybrid supercapacitors using fractional-order adaptive unscented Kalman filter," *Energy*, vol. 294, p. 130942, May, doi: https://doi.org/10.1016/j.energy.2024.130942.

[10] **H. Li, W. He, X. Fang, and S. Li**, "State-of-Charge Estimation of Supercapacitors: A Switched Systems Approach," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 3, pp. 5987–5996, Sep. 2024, doi: https://doi.org/10.1109/tte.2023.3328141.

[11] **P. Saha, S. Dey, and Munmun Khanra**, "Modeling and State-of-Charge Estimation of Supercapacitor Considering Leakage Effect," *IEEE transactions on industrial electronics*, vol. 67, no. 1, pp. 350–357, Jan. 2020, doi: https://doi.org/10.1109/tie.2019.2897506.

[12] **V. Bianchi, F. Canzanella, S. Ali, I. De Munari, L. Ciani, and G. Patrizi**, "A single-point EIS measurement for SOC estimation of supercapacitor," *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, May 2025, doi: https://doi.org/10.1109/i2mtc62753.2025.11078988.

[13] **S. Gow, M. Niranjan, S. Kanza, and J. G. Frey**, "A review of reinforcement learning in chemistry," *Digital Discovery*, vol. 1, no. 5, pp. 551–567, 2022, doi: https://doi.org/10.1039/d2dd00047d.

[14] **A. C. Erdem, V. Mert, M. A. Külünk, M. S. Oruç, T. Altun, and D. A. Kocabas**, "State of charge estimation for a UAV battery via Adaptive Extended Kalman Filter and Deep Reinforcement Learning," Engineering Science and Technology, an International Journal, vol. 77, Art. no. 102369, 2026, doi: 10.1016/j.jestch.2026.102369.

[15] **I. Saba, M. Tariq, M. Ullah, and H. Vincent Poor**, "Deep reinforcement learning based state of charge estimation and management of electric vehicle batteries," *IET smart grid*, vol. 6, no. 4, pp. 422–431, Jun. 2023, doi: https://doi.org/10.1049/stg2.12110.

[16] **Emirhan Ilhan, A. B. Koc, and S. S. Kozat**, "Exploiting residual errors in nonlinear online prediction," *Machine Learning*, vol. 113, no. 9, pp. 6065–6091, May 2024, doi: https://doi.org/10.1007/s10994-024 06554-7.

[17] **S. Ali, G. Patrizi, L. Ciani, I. De Munari, and V. Bianchi**, "Signal-to-Noise Ratio Analysis for a Reinforcement Learning based State of Charge Estimation of a Hybrid Supercapacitor," accepted for presentation at the *2026 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2026.

[18] **V. Bianchi, M. Stighezza, A. Toscani, G. Chiorboli, and I. De Munari**, "An Improved Method Based on Support Vector Regression With Application Independent Training for State of Charge Estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023, doi: https://doi.org/10.1109/tim.2023.3306816.

[19] **G. Patrizi, S. Ali, L. Ciani, F. Canzanella, I. De Munari and V. Bianchi**, "A Fast Approach for In Situ SoC Estimation of a Hybrid Supercapacitor Based on GPR and EIS Measurements" *IEEE Transactions on Instrumentation and Measurement*, vol. 75, pp. 1-11, 2026, doi: 10.1109/TIM.2026.3667312.