iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** XI **Month of publication:** November 2026

DOI: https://doi.org/10.22214/ijraset.2026.85888

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

Uncertainty Quantification in Lithium-Ion Battery State-of-Charge Estimation for Electric Vehicles: A Systematic Review

Swaruop Narain Shukla[1], Dileep Kumar[2], Ankita Tiwari[3]
[1, 3]*Department of Electrical Engineering, VSGOI Unnao, India*
[2]*Department of Applied Science and Humanities, VSGOI Unnao, India*

Abstract: ***State of charge (SOC) estimates in electric vehicles are usually reported as single numbers, although the battery management system needs to know how far each estimate can be trusted. This systematic review, reported under PRISMA 2020, examines how uncertainty is quantified in lithium-ion SOC estimation. Studies were identified through eleven concept-specific searches across publisher platforms and repositories, and 20 of 32 assessed studies, published from 2013 to 2026, met the eligibility criteria. They fall into five families: probabilistic filtering and uncertainty propagation, Gaussian process regression, Gaussian processes embedded in Bayesian filters, interval observers, and deep-learning methods based on Monte Carlo dropout, quantile loss, conformal calibration, evidential output and model disagreement. Deep-learning approaches account for eight studies, all published from 2023 onward. Model-based methods make parameter and measurement uncertainty explicit, whereas most data-driven intervals capture only data noise. Sixteen studies display an uncertainty output without testing its coverage; one reports near-nominal coverage with interval width and calibration error, and one tests behaviour along realistic ageing paths. No study reports interval coverage per temperature or drive cycle, and none reports cost on embedded hardware. We propose shared benchmarks with condition-stratified calibration metrics, coverage methods suited to autocorrelated and drifting data, separate reporting of aleatoric and epistemic uncertainty, pack-level propagation, and measured embedded performance as priorities for trustworthy SOC estimation.***

Keywords: battery management system, calibration, conformal prediction, Gaussian process, Bayesian deep learning, electric vehicle.

I. INTRODUCTION

An electric vehicle (EV) driver reads the state of charge (SOC) the way a petrol driver reads a fuel gauge, and the battery management system (BMS) uses the same quantity to decide how hard the pack may be charged or discharged [1]. Because SOC cannot be measured directly inside a lithium-ion cell, every BMS infers it from terminal voltage, current and temperature through some combination of Coulomb counting, open-circuit voltage lookup, equivalent circuit or electrochemical models, and data-driven regressors [1], [2]. An error in that inference does not stay on the dashboard. It shifts the usable energy window, can push cells toward over-charge or over-discharge, and changes the range a driver is promised [1].

Most published SOC estimators report a single number per time step and judge it by root mean square error (RMSE) or mean absolute error (MAE) on a held-out drive cycle. Kalman-filter estimators are a partial exception: since the extended Kalman filter (EKF) formulation of Plett [3], recursive Bayesian filters have carried an error covariance alongside the state, and sigma-point variants were introduced partly because the linearisation assumptions behind the EKF are not met exactly in battery applications [4]. Data-driven estimators moved in the opposite direction. Recurrent networks such as the long short-term memory (LSTM) model of Chemali et al. [5] map measured signals to SOC with low error and without an explicit cell model, and later surveys identify machine learning as a leading route to fast state prediction [6]; yet a standard network of this kind returns no statement of how far its output can be trusted. Gaussian process regression (GPR) closed part of that gap by attaching a predictive variance to each SOC estimate [7]. This matters because a point estimate with a small average error can still be badly wrong in the conditions a vehicle actually meets: a temperature the model never saw, an aged cell, a sensor offset. Neural regressors are known to give overconfident outputs when they leave the training distribution [8], and the machine learning community now separates aleatoric uncertainty, which comes from noise in the data, from epistemic uncertainty, which comes from limited knowledge of the model and shrinks with more data [9], [10]. A BMS that knows its own uncertainty can widen safety margins, request a corrective rest-voltage measurement, or flag a cell for inspection. One that does not must apply the same fixed margin everywhere.

©IJRASET: All Rights are Reserved

Existing reviews leave this question open for SOC. Broad SOC surveys organise the field by estimator family and report accuracy, with uncertainty handled at most as a by-product of GPR [1], [2], [11]. The recent review of probabilistic machine learning by Thelen et al. [12] treats uncertainty in depth, but its scope is state of health (SOH) and lifetime prognostics, with SOC estimation mentioned only as background. Tutorials on uncertainty quantification (UQ) for prognostics and health management likewise draw their case studies from degradation problems rather than SOC [13]. As a result, there is no systematic account of which UQ methods have been applied to SOC estimation, which sources of uncertainty they capture, how their intervals are evaluated, and whether those intervals hold up under EV operating conditions.

This review addresses that gap through a systematic search and synthesis conducted under the PRISMA 2020 reporting guideline [14]. It is organised around four research questions (RQs):

- RQ1. Which UQ methods have been applied to lithium-ion SOC estimation, and how has their use changed over time?
- RQ2. Which sources of uncertainty (measurement, model, parameter, and data or distribution shift) does each method represent, and which does it ignore?
- RQ3. Which datasets, test protocols and metrics are used to evaluate the quality of SOC uncertainty estimates, as opposed to point accuracy?
- RQ4. What evidence exists on calibration under temperature change, ageing and unseen drive cycles, and on the computational cost of UQ on embedded BMS hardware?

The contributions of the review are as follows:

1) A PRISMA 2020-reported search, screening and preliminary quality appraisal of studies that report uncertainty for SOC estimates in lithium-ion cells and packs.

2) A taxonomy that places Bayesian filters, Gaussian processes, Bayesian and approximate Bayesian neural networks, ensembles, evidential models, quantile models and conformal prediction on a common footing, mapped to the uncertainty sources each one covers.

3) A synthesis of evaluation practice, showing which studies test interval coverage and calibration and which report only point error.

4) A research agenda for uncertainty-aware SOC estimation in EVs, covering calibration under distribution shift, pack-level propagation and on-board deployment.

Section II gives background on SOC estimation and UQ. Section III describes the review protocol. Section IV reports the results of the synthesis, Section V discusses open problems, and Section VI concludes.

II. BACKGROUND

A. State of Charge and Its Estimation

SOC expresses the charge remaining in a cell as a fraction of its usable capacity. In discrete time with sampling interval Δt, the Coulomb-counting definition used by most estimators is

$$z_{k+1} = z_k - \frac{\eta \Delta t}{Q} i_k \qquad (1)$$

where z_k is SOC at step k, i_k is the measured current (positive on discharge), η is the Coulombic efficiency and Q is the cell capacity in ampere-seconds. The equation looks exact, but each symbol on its right side is uncertain: i_k carries sensor noise and offset, η and Q drift with temperature and ageing, and the initial value z_0 is rarely known precisely. Open-loop integration therefore accumulates error, which is why practical estimators correct it with a voltage-based observation [1], [2].

Published estimators fall into three broad families [1], [2]. Model-based methods pair an equivalent circuit or electrochemical model with an observer, most often a Kalman-type filter; Plett's cell model, for example, includes open-circuit voltage, ohmic loss, polarisation time constants, hysteresis and temperature terms [15]. Data-driven methods learn the mapping from measured signals to SOC directly, using support vector machines, GPR or deep networks [5], [6], [7]. Hybrid methods combine the two, for instance by using a network to estimate model parameters or a filter to smooth network outputs.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Where Uncertainty Enters an SOC Estimate

It is useful to separate four sources, because UQ methods differ in which ones they represent.

1) Measurement uncertainty. Noise, quantisation and bias in current, voltage and temperature sensors. Current offset is especially harmful because Coulomb counting integrates it.
2) Model-form uncertainty. The gap between the assumed cell model, or the learned regressor, and the real electrochemistry, including unmodelled hysteresis and relaxation dynamics.
3) Parameter uncertainty. Spread in capacity, resistance and time constants from cell-to-cell variation, temperature dependence and ageing.
4) Distribution shift. Operating conditions at deployment that differ from those used for training or identification: new temperatures, drive cycles, chemistries or states of health.

In machine learning terms, the first source is largely aleatoric, while the last three are mainly epistemic and can in principle be reduced with better models or more representative data [9], [10]. The distinction has practical weight. Aleatoric uncertainty sets a floor that no estimator can beat with the same sensors, whereas a large epistemic component signals that the estimate should not be trusted until more information arrives.

C. Families of Uncertainty Quantification Methods

The UQ methods that appear in battery state estimation can be grouped into six families. Fig. 1 maps each family to the uncertainty sources of Section II-B that it represents explicitly.

1) Recursive Bayesian filters. The EKF, sigma-point (unscented) Kalman filters and particle filters propagate a state distribution through the cell model and update it with each voltage measurement [3], [4]. Their covariance reflects the process and measurement noise that the designer specifies, so its reliability depends on how well those noise models match reality.
2) Gaussian process regression. GPR places a prior over functions and returns a Gaussian predictive distribution whose variance grows away from the training data [7], [13].
3) Bayesian and approximate Bayesian neural networks. Variational inference places distributions on network weights; Monte Carlo (MC) dropout approximates this by keeping dropout active at test time and averaging several stochastic forward passes [16], [17].
4) Deep ensembles. Several networks trained from different initialisations are combined, and their disagreement serves as an estimate of epistemic uncertainty; when each member also predicts a variance, aleatoric uncertainty is captured as well [8].
5) Evidential and distributional networks. A single deterministic network outputs the parameters of a higher-order distribution, such as the Normal-Inverse-Gamma prior of deep evidential regression, giving both uncertainty types without sampling [18].
6) Distribution-free methods. Quantile regression and conformal prediction produce intervals directly; split conformal prediction wraps any trained model and guarantees marginal coverage at a chosen level, provided calibration and test data are exchangeable [19].

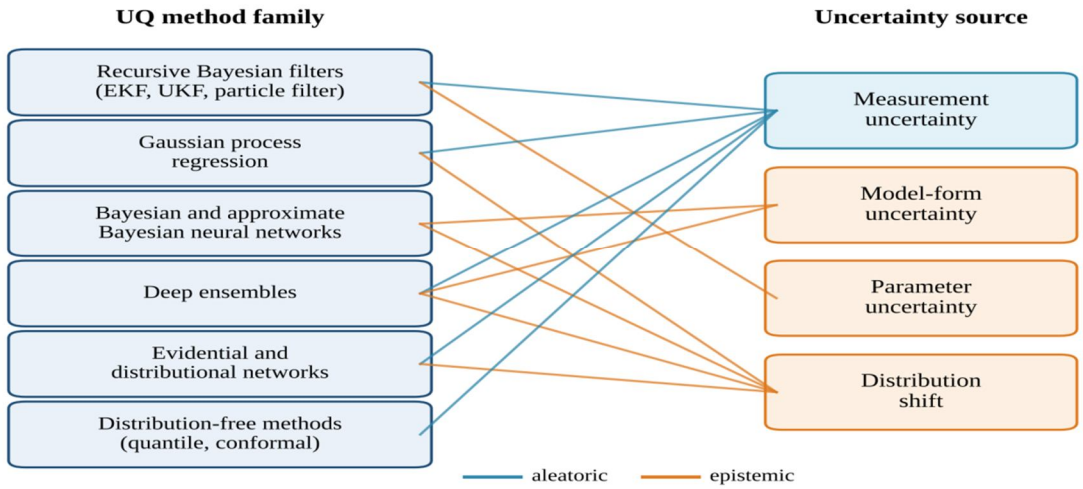


Fig. 1 UQ method families and the uncertainty sources each represents explicitly (blue: aleatoric; orange: epistemic), as described in Sections II-B and II-C.

The exchangeability condition deserves attention for SOC. Consecutive samples within a drive cycle are strongly autocorrelated, and an EV pack moves through temperatures and ageing states that a calibration set may not contain. A coverage guarantee proven for exchangeable data may therefore not transfer to a vehicle without adaptation, a point we return to in Section V.

D. Evaluating an Uncertainty Estimate

An interval or predictive distribution is useful only if it is both calibrated and informative. For a test set of N points with true SOC y_i and a nominal $(1 - \alpha)$ interval $[L_i, U_i]$, two widely reported metrics are the prediction interval coverage probability (PICP) and the mean prediction interval width (MPIW):

$$\mathrm{PICP} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[L_i \le y_i \le U_i], \qquad \mathrm{MPIW} = \frac{1}{N}\sum_{i=1}^{N} (U_i - L_i) \tag{2}$$

A PICP close to $1 - \alpha$ with a small MPIW is the goal; a wide interval achieves coverage trivially and is of little use to a BMS. Strictly proper scoring rules such as the negative log-likelihood (NLL) and the continuous ranked probability score (CRPS) reward calibration and sharpness together, and so are preferred for comparing probabilistic estimators [20]. Calibration across all confidence levels can be inspected with a reliability diagram that plots observed against expected coverage; Kuleshov et al. showed that Bayesian regressors are often miscalibrated in this sense and proposed a post-hoc recalibration step [21]. We use these metrics in Section IV to judge how thoroughly each study evaluated its uncertainty estimates.

III.REVIEW METHODOLOGY

The review follows the PRISMA 2020 statement for reporting [14] and the planning, conducting and reporting stages recommended for systematic literature reviews in computing and engineering [22]. The research questions in Section I, the search concepts and the eligibility criteria in Table I were fixed before screening began. The protocol was not registered.

A. Information Sources and Search Strategy

Studies were identified on 19 November 2026 through a web search engine that indexes publisher platforms (ScienceDirect, IEEE Xplore, SpringerLink, Wiley Online Library, MDPI, Nature Portfolio), institutional repositories and indexing services, rather than through export from a single bibliographic database. Eleven discovery queries were run. Each paired "state of charge" and "lithium-ion" with one uncertainty concept: Gaussian process regression, Bayesian neural network, Monte Carlo dropout, deep ensemble with prediction interval, conformal prediction, evidential deep learning, set-membership or interval observer, particle filter with probability density or confidence interval, quantile regression, Gaussian process Kalman filter, and variational Bayesian recurrent networks. For every candidate, the bibliographic record was then retrieved from the publisher page or from an indexing or institutional repository record. The Boolean string below is the equivalent Scopus query, provided so that the identification step can be replicated at database level:

```
TITLE-ABS-KEY ( ( "lithium-ion" OR "li-ion" OR "lithium ion" ) AND battery )
AND TITLE-ABS-KEY ( "state of charge" OR "state-of-charge" OR SOC )
AND TITLE-ABS-KEY ( "uncertainty quantification" OR "predictive uncertainty"
  OR "confidence interval" OR "prediction interval" OR "credible interval"
  OR probabilistic OR bayesian OR "gaussian process" OR "monte carlo dropout"
  OR ensemble OR evidential OR "conformal prediction" OR "quantile regression"
  OR "error covariance" OR "error bound" )
AND PUBYEAR > 2009 AND LANGUAGE ( english )
AND DOCTYPE ( ar OR cp OR re )
```

Backward citation searching was applied whenever a candidate's references or citing records pointed to another eligible study. Because identification relied on a search engine, the number of records at the identification stage is not reproducible in the strict PRISMA sense. Section IV-A therefore reports the outcome for every record that reached eligibility assessment, and the lack of a database-level count is treated as a limitation in Section V-E.

B. Eligibility Criteria

TABLE I
ELIGIBILITY CRITERIA

Criterion	Inclusion	Exclusion
Population	Lithium-ion cells, modules or packs of any chemistry	Lead-acid, nickel-metal hydride, sodium-ion, supercapacitors
Task	SOC estimation, alone or jointly with other states	SOH, remaining useful life or power estimation without an SOC output
Uncertainty	Reports a quantitative uncertainty measure for the SOC estimate (variance, covariance, interval, distribution or error bound)	Point estimates only, or uncertainty mentioned without being computed
Evidence	Experimental or simulation results with a described dataset	Purely conceptual papers, patents, editorials
Publication	Peer-reviewed journal article or indexed conference paper in English, 2010 onward	Abstracts only, theses, non-English texts, retracted papers

Studies of Kalman-type and particle filters were included only when the state distribution, covariance or a derived interval was reported or evaluated as an output. Filters used purely as point estimators were excluded, because almost every filter-based SOC paper computes a covariance internally without assessing it.

C. Selection Process

Titles and search snippets were screened first, then abstracts and, where openly available, full texts were checked against the criteria in Section III-B. Screening and extraction were performed by one reviewer with AI assistance (Claude, Anthropic). [DETAIL_TODO: independent check by a second author, with the agreement statistic, before submission]

D. Data Extraction

A structured form was applied to each included study. It recorded: bibliographic details; cell chemistry and format; dataset and whether it is public; operating conditions used for training and testing; base SOC estimator; UQ family (Section II-C); uncertainty sources represented (Section II-B); uncertainty metrics reported; whether shifted conditions such as ageing were tested; and any statement on computational cost. Fields that could not be read from the abstract or openly available text were recorded as not reported. Public datasets were identified against the catalogue compiled by dos Reis et al. [23].

E. Quality Appraisal

Because no established appraisal tool fits SOC uncertainty studies, a checklist of eight items was adapted from the extraction fields above. Each item was scored 1 (yes), 0.5 (partly) or 0 (no):

1) The dataset is described well enough to be reproduced, or is public.
2) Training and test data are separated by drive cycle, temperature or cell, not by random sampling of time steps.
3) A baseline point estimator is reported for comparison.
4) At least one uncertainty metric beyond point error is reported.
5) Coverage or calibration is checked against the nominal level.
6) Performance is tested under at least one shifted condition (temperature, ageing or unseen cycle).
7) Computational cost or embedded feasibility is reported.
8) Code or trained models are available.

Because several included studies could be read only at abstract level, this appraisal is preliminary. Items that could not be confirmed were scored 0, and no study was excluded on quality grounds; Table V reports items 4 to 6, which bear most directly on the research questions, and the strength of each conclusion in Section V is matched to that evidence.

F. Synthesis

The included studies differ widely in chemistry, data, protocol and metric, so a quantitative meta-analysis of interval quality was not attempted. Instead, results are tabulated by UQ family and synthesised narratively against RQ1 to RQ4. Where several studies used the same public dataset and test protocol, their reported metrics are placed side by side, with protocol differences noted.

IV. RESULTS

A. Study Selection

Thirty-two records reached eligibility assessment, and 20 studies published between 2013 and 2026 met all criteria. The twelve exclusions fell into four groups (Table II), and Fig. 2 summarises the selection process:

TABLE II
REASONS FOR EXCLUDING ASSESSED STUDIES

Reason for exclusion	**Records**	**Examples**
Uncertainty quantified for SOH or RUL, not SOC	3	Bayesian deep learning for SOH; MC dropout for RUL; conformal prediction for SOH
SOC estimated, but no uncertainty output computed or reported	4	Ensemble networks used to improve point accuracy; GPR used as a point regressor; Kalman-GPR hybrid reporting only relative error
Task outside SOC estimation	2	Particle-filter discharge-time prognosis risk measures; physics-informed discharge prognosis
Bibliographic record could not be verified	3	Conference and journal records whose venue or year could not be confirmed

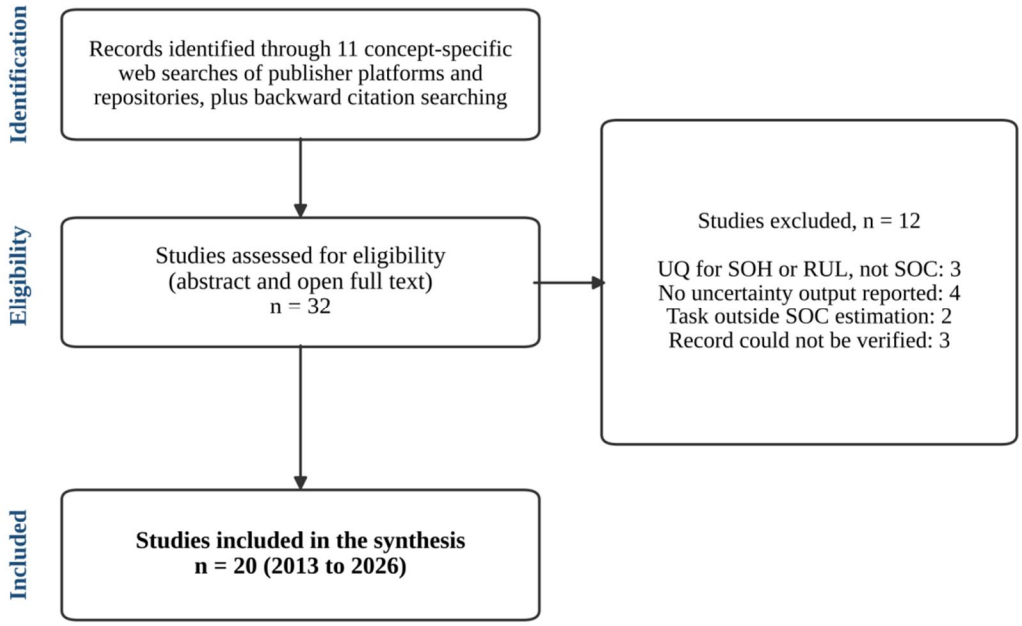


Fig. 2 Selection of studies. Identification used concept-specific web searches, so no database-level record count is shown.

B. Methods Applied and Their Trend (RQ1)

Table III lists the included studies by UQ family. Model-based probabilistic filtering is the oldest strand: Schwunk et al. [29] used a particle filter for lithium iron phosphate (LFP) cells because their flat, charge-direction-dependent open-circuit voltage makes a Gaussian mean-and-variance description fail, and Pola et al. [30] used particle filtering to estimate SOC and model parameters in real time with a full probability density. Gaussian process regression entered in 2016 with an online estimator whose kernel choice and uncertainty were studied explicitly [24], followed by its regular and recurrent extension [7], a pack-level autoregressive variant

[25] and a deep recurrent kernel [26]. Two groups moved Gaussian processes inside Bayesian filters, using them as learned process and measurement models so that model variability enters the prediction and update steps [27], [28].

Deep-learning UQ for SOC is recent (Fig. 3). All eight deep-learning studies appeared from 2023 onward, and eight of the 20 included studies were published in 2025 or 2026. Within that group no single method dominates: quantile-loss networks [36], [37], conformalized quantile regression [38], evidential regression [39], [40], MC dropout combined with Monte Carlo propagation of measurement noise [35], a comparison of NGBoost, MC dropout and GPR on ultrasonic features [41], and model disagreement between an LSTM and a GPR used as an error indicator under ageing [42].

TABLE III
INCLUDED STUDIES BY UQ FAMILY

Ref.	Year	UQ family	Base estimator or model	Uncertainty output
[29]	2013	Particle filter	Cell model, LFP chemistry	Full SOC probability density
[30]	2015	Particle filter	Empirical state-space model	SOC density and 95% interval
[24]	2016	Gaussian process regression	GPR on voltage, current, temperature	Predictive variance
[7]	2018	Gaussian process regression	Regular and recurrent GPR	Predictive variance
[27]	2019	GP inside Bayesian filter	GP-based Bayes filters	Filtered SOC distribution
[25]	2020	Gaussian process regression	Autoregressive GPR, battery pack	Confidence interval
[26]	2020	Gaussian process regression	GPR with GRU-based kernel	Predictive variance
[33]	2020	Interval observer	Electrical-thermal model, parallel cells	Guaranteed SOC bounds
[31]	2021	Uncertainty propagation with EKF	EKF with GP-modelled parameters	SOC distribution over a cell batch
[28]	2023	GP inside Bayesian filter	GP-UKF and GP-PF	Filtered SOC distribution
[35]	2023	MC dropout and Monte Carlo simulation	2D CNN on impedance spectra	Classification uncertainty
[36]	2024	Quantile regression	Hybrid network with attention	95% interval
[34]	2025	Interval observer	Second-order equivalent circuit	Dynamic error bounds
[37]	2025	Quantile regression	CNN-LSTM with attention	Confidence interval
[40]	2025	Evidential regression	GRU	Aleatoric and epistemic variance
[42]	2025	Model disagreement	LSTM and GPR	Disagreement index as error proxy
[32]	2026	Particle filter	Residual-augmented particle filter	Weighted-quantile interval
[38]	2026	Conformalized quantile regression	Transformer-type backbone	Calibrated interval
[39]	2026	Evidential regression	Temporal-spectral fusion network	Normal-Inverse-Gamma uncertainty
[41]	2026	Several (NGBoost, MC dropout, GPR)	Ultrasonic phase and amplitude features	Prediction interval

TABLE IV
NUMBER OF INCLUDED STUDIES BY FAMILY AND PERIOD

UQ family	2013 to 2019	2020 to 2022	2023 to 2026	Total
Probabilistic filtering and propagation	2	1	1	4
Gaussian process regression	2	2	0	4
GP inside Bayesian filter	1	0	1	2
Interval observer	0	1	1	2
Deep-learning UQ (dropout, quantile, conformal, evidential, disagreement, mixed)	0	0	8	8
Total	**5**	**4**	**11**	**20**

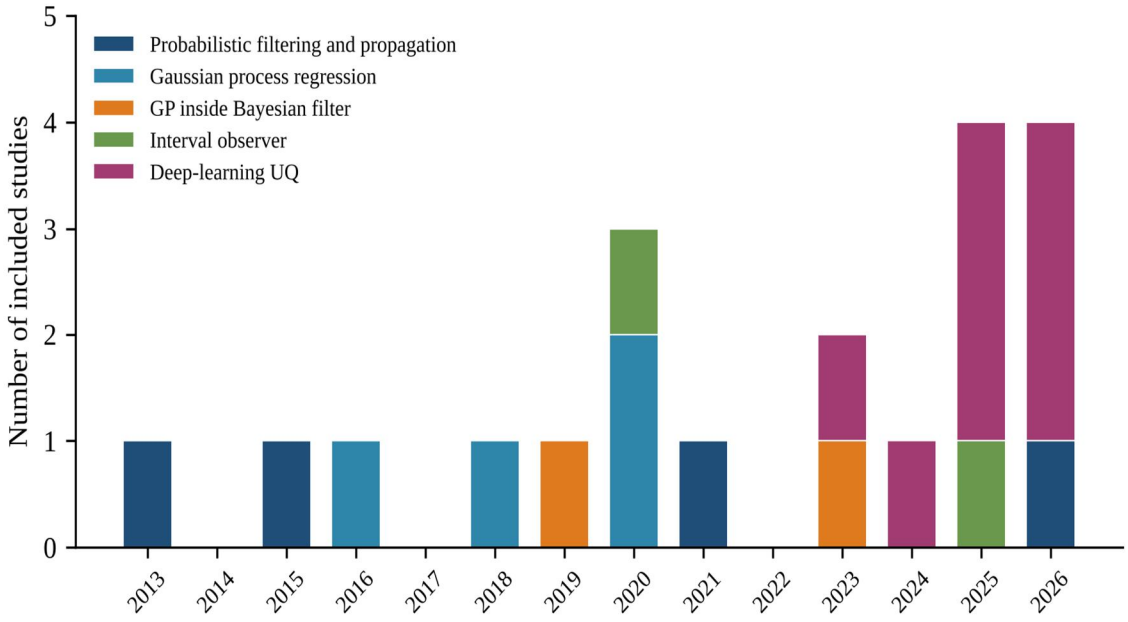


Fig. 3 Included studies per publication year, coloured by UQ family (data from Table III).

C. Uncertainty Sources Represented (RQ2)

The families divide sharply by what they treat as uncertain. Model-based studies make parameter and measurement uncertainty explicit. Dahmardeh and Xi [31] modelled equivalent-circuit parameters as Gaussian processes over SOC to represent cell-to-cell manufacturing variation, so that one estimator gives a reliable SOC distribution for a batch of similar cells without characterising each cell. The interval observers treat parameters, initial states and noise as unknown but bounded: Zhang et al. [33] mapped such bounds to a feasible SOC set for coupled parallel cells, and Wu et al. [34] derived dynamic error bounds that remain valid under parametric perturbation and noise. The LFP particle filter [29] addresses model-form ambiguity, since a flat, hysteretic voltage curve admits more than one SOC for a given measurement.

Most deep-learning studies, by contrast, represent the spread of the data around the regression output. Quantile and conformalized-quantile networks [36], [37], [38] produce intervals that reflect noise in the training targets but carry no explicit term for model ignorance. The evidential models [39], [40] are the exception within this group, because the Normal-Inverse-Gamma output separates the two components in a single forward pass. Only one study targeted distribution shift directly: Xue et al. [42] quantified how LSTM and GPR estimators degrade along complex ageing paths and used model disagreement to decide when to retrain.

D. Evaluation of Uncertainty Estimates (RQ3)

Evaluation practice is the weakest part of the corpus. Of the 20 studies, only four report a quantitative check that the uncertainty is informative or calibrated, and each uses a different measure (Table V, Fig. 4). Soon and Soon [38] give the most complete set: coverage close to the 95% target, a normalised mean interval width of 0.1452, a miscalibration error of 0.0194 and a Winkler score of 13.636, compared with the same backbone before conformal calibration. Wu et al. [39] judged their evidential uncertainty by its ability to flag large errors and report that it identified 94% of critical estimation deviations. Xue et al. [42] report that their disagreement index correlates with the true SOC error at 0.91 and that the closed-loop scheme kept average error below 1.5% with four retrainings on the WLTC profile. The evidential GRU study [40] compared its single-pass uncertainty with ensembles and found comparable performance on in-distribution data.

The remaining studies show intervals, densities or bounds, usually as shaded bands around a test trajectory, without testing their coverage. This is true of every Gaussian process study in the set [7], [24], [25], [26], although the autoregressive pack model [25] does note that its confidence interval narrows relative to the regular GPR. Interval-observer papers prove inclusion of the true state analytically [33], [34], which is a stronger guarantee than empirical coverage but holds only if the assumed bounds are correct.

TABLE V
APPRAISAL ITEMS 4 TO 6, AS REPORTED IN THE ABSTRACT OR OPEN TEXT

Ref.	Item 4: uncertainty metric beyond point error	Item 5: coverage or calibration checked	Item 6: shifted condition tested
[38]	Yes (PICP, NMPIW, MCE, Winkler)	Yes	Not reported
[39]	Yes (deviation detection rate)	Partly	Partly (transferability examined)
[42]	Yes (correlation with true error)	Partly	Yes (ageing paths)
[40]	Yes (comparison with ensembles)	Partly	Partly (in-distribution only reported)
[30]	Partly (95% interval shown)	Not reported	Not reported
[33], [34]	Partly (analytic inclusion)	Partly (by proof)	Not reported
[7], [24], [25], [26], [27], [28], [29], [31], [32], [35], [36], [37], [41]	Partly (interval, variance or density shown)	Not reported	Not reported

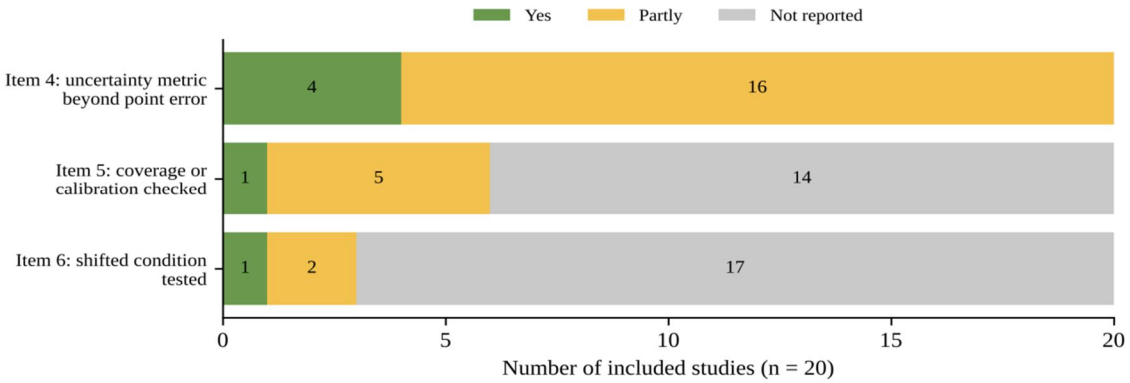


Fig. 4 Appraisal items 4 to 6 across the 20 included studies (data from Table V).

Public data appear in only some studies. Several deep-learning SOC works in the wider literature rely on the Panasonic 18650PF drive-cycle data at five temperatures released by Kollmeyer [43], whereas included studies also used an LFP cell from A123 [29], Samsung ICR18650-26J impedance spectra [35] and in-house ultrasonic measurements [41]. With different cells, protocols and metrics, a quantitative comparison of interval quality across families is not possible from the published evidence.

E. Robustness to Shift and Computational Cost (RQ4)

Evidence on behaviour outside the training distribution is limited to three studies [39], [40], [42], and only one of them [42] tests degradation along realistic ageing trajectories. No included study reported interval coverage separately for unseen temperatures or drive cycles. Computational cost is discussed qualitatively: evidential regression avoids the repeated forward passes of ensembles and dropout sampling [40], quantile-loss training avoids the matrix inversion of Bayesian methods [36], and the parallel-cell interval observer is independent of the number of cells in the module [33]. No included study reported memory use, latency or deployment on a production BMS microcontroller.

V. DISCUSSION

A. What Uncertainty-aware SOC Estimation Delivers Today

The 20 studies show that SOC uncertainty can be produced by almost any estimator family, from particle filters to transformers. What the corpus does not yet show is that those uncertainties are trustworthy in the conditions an EV meets. The model-based strand has the clearer link between uncertainty and physics: when parameters vary from cell to cell [31], or when the voltage curve is flat and hysteretic [29], the method states which quantity is uncertain and propagates it. Its weakness is that the result is only as good as the assumed noise model or bounds, and none of the filter studies checked whether their reported spread matched observed errors.

The data-driven strand inverts the trade-off. Quantile and conformal networks need no noise model and can be calibrated empirically, and the one study that did so reached near-nominal coverage [38]. Yet most of these intervals describe aleatoric spread only. A network trained at 25 °C can return a narrow quantile band at −10 °C because nothing in the quantile loss penalises confidence far from the training data. Evidential output [39], [40] and model disagreement [42] are the only data-driven mechanisms in the set that respond to ignorance as well as noise.

B. Calibration Is Rarely Tested

The most consistent finding is that interval quality is reported far less often than interval existence. Sixteen of 20 studies display an uncertainty output without measuring its coverage (Table V). Kuleshov et al. [21] showed in general regression tasks that Bayesian credible intervals are frequently miscalibrated, so a shaded band in a figure is not evidence that a 95% band contains the truth 95% of the time. For SOC this matters in both directions: an interval that is too narrow invites an unsafe discharge limit, while one that is too wide wastes usable energy through excessive margins. Proper scoring rules such as CRPS and NLL [20], reliability diagrams, and PICP with MPIW at several nominal levels should become standard alongside RMSE.

Conformal prediction offers a route to guaranteed coverage, but its guarantee assumes exchangeable calibration and test data [19]. SOC time series are autocorrelated, and operating temperature and ageing drift over the life of a pack. A calibration set drawn from the same drive cycles as the training data will therefore overstate coverage on the road. Time-series and adaptive conformal variants, calibration sets stratified by temperature and state of health, and reporting of coverage per condition are needed before a coverage claim can transfer to a vehicle.

C. From Cells to Packs

Only three studies addressed battery packs or batches of cells as the object of estimation [25], [31], [33]. EV packs contain hundreds of cells, and the pack-level quantity a driver sees depends on the weakest series group. Propagating per-cell uncertainty through series and parallel connections, as the interval observer does for parallel cells [33], is largely unexplored for probabilistic and learned estimators. Cell-to-cell variation is also a natural source of epistemic uncertainty for data-driven models trained on one or a few cells.

D. Deployment on a BMS

No study reported latency, memory or energy use on embedded hardware. This leaves an open question for methods that need repeated inference, such as MC dropout and ensembles, or matrix operations that grow with the training set, such as exact GPR. Single-pass methods (evidential, quantile, conformal with a fixed calibration quantile) and set-based observers whose cost does not grow with the number of cells [33] appear better suited to embedded deployment, but this needs measured evidence rather than inference from method structure.

E. Limitations of This Review

Three limitations qualify these findings. First, studies were identified through a web search engine across publisher platforms rather than by exporting records from Scopus and Web of Science, so the identification count is not reproducible and relevant studies may have been missed; a database-level rerun of the string in Section III-A is required before the review can claim completeness. Second, screening and extraction were done by one reviewer with AI assistance, and several studies were assessed at abstract level only, which may understate the evaluation they actually report. Third, the field is moving quickly: eight of the included studies are from 2025 and 2026, and further work will have appeared after the search date.

F. Research Agenda

The synthesis points to six concrete directions:

1) Shared benchmark protocols. Public datasets with defined splits by temperature, drive cycle and state of health, scored with PICP, MPIW, CRPS and calibration error at several confidence levels.

2) Coverage under shift. Conformal and recalibration methods that remain valid for autocorrelated data and across temperature and ageing, with coverage reported per condition.

3) Separated uncertainty sources. Estimators that report aleatoric and epistemic components separately, so a BMS can tell sensor noise from model ignorance and respond differently.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

4) Hybrid physics and learning. Combining the explicit parameter and measurement uncertainty of filters [28], [31] with learned residual models, and calibrating the result empirically.
5) Pack-level propagation. Methods that carry cell-level uncertainty to pack SOC and usable energy, including cell-to-cell variability.
6) Embedded evidence. Reported latency, memory and energy on automotive microcontrollers, and closed-loop tests that show how uncertainty changes BMS decisions such as power limits or recalibration requests.

VI. CONCLUSION

This review examined how uncertainty is quantified in lithium-ion SOC estimation for electric vehicles. Twenty studies from 2013 to 2026 met the inclusion criteria, spanning particle and Gaussian-process-based filters, Gaussian process regression, interval observers and, since 2023, deep-learning methods based on dropout, quantile loss, conformal calibration, evidential output and model disagreement. More than half of the studies appeared from 2023 onward.

The central finding is a gap between producing uncertainty and validating it. Sixteen of the 20 studies display an interval, variance, density or bound without testing whether it achieves its nominal coverage, and only one reported near-nominal coverage together with interval width and calibration error. Few studies addressed temperature or ageing shift, only three treated packs or cell batches, and none reported cost on embedded BMS hardware.

The review is limited by search-engine-based identification and single-reviewer screening, and a database-level search should confirm its coverage. The next step for the field is a shared benchmark that scores calibrated SOC intervals under realistic temperature, drive-cycle and ageing shift, so that uncertainty-aware estimators can be compared and trusted on the road.

VII. DECLARATIONS

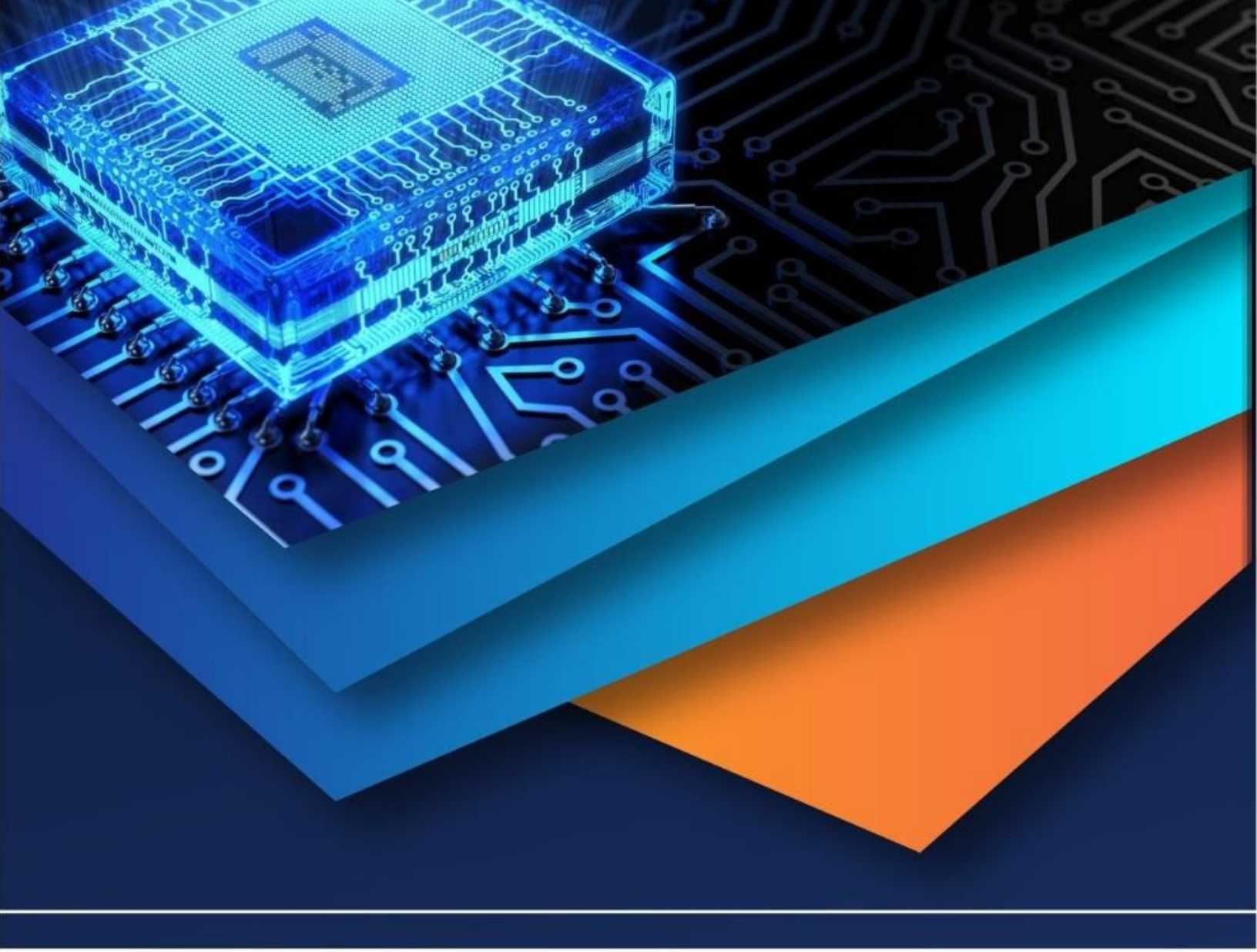
1) Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
2) Declaration of competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
3) Ethics statement. This review analyses published literature only and did not involve human participants or animals.

REFERENCES

[1] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 834-854, 2017, doi: 10.1016/j.rser.2017.05.001.
[2] D. N. T. How, M. A. Hannan, M. S. Hossain Lipu, and P. J. Ker, "State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review," *IEEE Access*, vol. 7, pp. 136116-136136, 2019, doi: 10.1109/ACCESS.2019.2942213.
[3] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background," *J. Power Sources*, vol. 134, no. 2, pp. 252-261, 2004, doi: 10.1016/j.jpowsour.2004.02.031.
[4] G. L. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1: Introduction and state estimation," *J. Power Sources*, vol. 161, no. 2, pp. 1356-1368, 2006, doi: 10.1016/j.jpowsour.2006.06.003.
[5] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730-6739, 2018, doi: 10.1109/TIE.2017.2787586.
[6] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161-170, 2020, doi: 10.1038/s42256-020-0156-7.
[7] G. O. Sahinoglu, M. Pajovic, Z. Sahinoglu, Y. Wang, P. V. Orlik, and T. Wada, "Battery state-of-charge estimation based on regular/recurrent Gaussian process regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4311-4321, 2018, doi: 10.1109/TIE.2017.2764869.
[8] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017)*, 2017, pp. 6402-6413.
[9] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017)*, 2017, pp. 5574-5584.
[10] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457-506, 2021, doi: 10.1007/s10994-021-05946-3.
[11] Y. Wu, D. Bai, [DETAIL_TODO: remaining authors], and F. Yang, "Advancements in the estimation of the state of charge of lithium-ion battery: A comprehensive review of traditional and deep learning approaches," *J. Mater. Inf.*, vol. 5, no. 2, Art. no. 18, 2025, doi: 10.20517/jmi.2024.84.
[12] A. Thelen, X. Huan, N. Paulson, S. Onori, Z. Hu, and C. Hu, "Probabilistic machine learning for battery health diagnostics and prognostics: Review and perspectives," *npj Mater. Sustain.*, vol. 2, Art. no. 14, 2024, doi: 10.1038/s44296-024-00011-1.
[13] V. Nemani, L. Biggio, X. Huan, Z. Hu, O. Fink, A. Tran, Y. Wang, X. Zhang, and C. Hu, "Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial," *Mech. Syst. Signal Process.*, vol. 205, Art. no. 110796, 2023, doi: 10.1016/j.ymssp.2023.110796.
[14] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.

[15] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification," *J. Power Sources*, vol. 134, no. 2, pp. 262-276, 2004, doi: 10.1016/j.jpowsour.2004.02.032.
[16] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, PMLR, vol. 48, 2016, pp. 1050-1059.
[17] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243-297, 2021, doi: 10.1016/j.inffus.2021.05.008.
[18] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," in *Adv. Neural Inf. Process. Syst. 33 (NeurIPS 2020)*, 2020, [DETAIL_TODO: pages].
[19] A. N. Angelopoulos and S. Bates, "Conformal prediction: A gentle introduction," *Found. Trends Mach. Learn.*, vol. 16, no. 4, pp. 494-591, 2023, doi: 10.1561/2200000101.
[20] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Statist. Assoc.*, vol. 102, no. 477, pp. 359-378, 2007, doi: 10.1198/016214506000001437.
[21] V. Kuleshov, N. Fenner, and S. Ermon, "Accurate uncertainties for deep learning using calibrated regression," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, PMLR, vol. 80, 2018, pp. 2796-2804.
[22] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering: A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7-15, 2009, doi: 10.1016/j.infsof.2008.09.009.
[23] G. dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy AI*, vol. 5, Art. no. 100081, 2021, doi: 10.1016/j.egyai.2021.100081.
[24] G. Ozcan, M. Pajovic, Z. Sahinoglu, Y. Wang, P. V. Orlik, and T. Wada, "Online state of charge estimation for lithium-ion batteries using Gaussian process regression," in *Proc. IECON 2016, 42nd Annu. Conf. IEEE Ind. Electron. Soc.*, 2016, pp. 998-1003, doi: 10.1109/IECON.2016.7793002.
[25] Z. Deng, X. Hu, X. Lin, Y. Che, L. Xu, and W. Guo, "Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression," *Energy*, vol. 205, Art. no. 118000, 2020, doi: 10.1016/j.energy.2020.118000.
[26] [DETAIL_TODO: authors], "State of charge estimation for lithium-ion battery based on Gaussian process regression with deep recurrent kernel," *Int. J. Electr. Power Energy Syst.*, vol. 124, Art. no. 106369, 2020, doi: 10.1016/j.ijepes.2020.106369.
[27] L. O. Avila, M. L. Errecalde, F. M. Serra, and E. C. Martínez, "State of charge monitoring of Li-ion batteries for electric vehicles using GP filtering," *J. Energy Storage*, vol. 25, Art. no. 100837, 2019, doi: 10.1016/j.est.2019.100837.
[28] K.-J. Lee, W.-H. Lee, and K.-K. K. Kim, "Battery state-of-charge estimation using data-driven Gaussian process Kalman filters," *J. Energy Storage*, vol. [DETAIL_TODO], Art. no. 108392, 2023, doi: 10.1016/j.est.2023.108392.
[29] S. Schwunk *et al.* [DETAIL_TODO: full author list], "Particle filter for state of charge and state of health estimation for lithium-iron phosphate batteries," *J. Power Sources*, vol. 239, pp. 705-710, 2013, doi: 10.1016/j.jpowsour.2012.10.058.
[30] D. A. Pola, H. F. Navarrete, M. E. Orchard, R. S. Rabie, M. A. Cerda, B. E. Olivares, J. F. Silva, P. A. Espinoza, and A. Pérez, "Particle-filtering-based discharge time prognosis for lithium-ion batteries with a statistical characterization of use profiles," *IEEE Trans. Reliab.*, [DETAIL_TODO: year, volume, issue, pages, DOI].
[31] M. Dahmardeh and Z. Xi, "Probabilistic state-of-charge estimation of lithium-ion batteries considering cell-to-cell variability due to manufacturing tolerance," *J. Energy Storage*, vol. [DETAIL_TODO], Art. no. 103204, 2021, doi: 10.1016/j.est.2021.103204.
[32] [DETAIL_TODO: authors], "Residual-allocation particle filtering for lithium-ion battery state-of-charge estimation under long dynamic operation," [DETAIL_TODO: journal, volume, article number, DOI; ScienceDirect PII S2590174526007440], 2026.
[33] D. Zhang, L. D. Couto, P. Gill, S. Benjamin, W. Zeng, and S. J. Moura, "Interval observer for SOC estimation in parallel-connected lithium-ion batteries," in *Proc. 2020 Amer. Control Conf. (ACC)*, 2020, pp. 1149-1154, doi: 10.23919/ACC45564.2020.9147468.
[34] C. Wu, Y. Wang, and Y. Xu, "Robust interval observer design with error bounds for lithium-ion battery state-of-charge estimation under hybrid uncertainties," [DETAIL_TODO: venue, volume, pages, DOI], 2025.
[35] E. Buchicchio, A. De Angelis, F. Santoni, and P. Carbone, "Uncertainty characterization of a CNN method for lithium-ion batteries state of charge estimation using EIS data," *Measurement*, vol. [DETAIL_TODO], Art. no. 113341, 2023, doi: 10.1016/j.measurement.2023.113341.
[36] H. Li, L. Fu, X. Long, L. Liu, and Z. Zeng, "A hybrid deep learning model for lithium-ion batteries state of charge estimation based on quantile regression and attention," *Energy*, vol. 294, Art. no. 130834, 2024, [DETAIL_TODO: DOI].
[37] Y. Zou, S. Wang, N. Hai, F. Blaabjerg, C. Fernandez, and W. Cao, "Enhanced quantile regression long short-term memory hybrid neural network for the state of charge point and interval estimation of lithium-ion batteries," *Energy*, vol. 332, Art. no. 137201, 2025, doi: 10.1016/j.energy.2025.137201.
[38] K. L. Soon and L. T. Soon, "Enhancing reliability in electrified transportation: A conformalized quantile regression framework for battery state-of-charge uncertainty quantification," *J. Power Sources*, vol. 666, Art. no. 239123, 2026, doi: 10.1016/j.jpowsour.2025.239123.
[39] Z. Wu, B. Xu, F. Wen, J. Zhao, and F. Shen, "An end-to-end dual-domain attention fusion network with evidential uncertainty for battery state of charge estimation," *J. Energy Storage*, vol. 148, Art. no. 119996, 2026, doi: 10.1016/j.est.2025.119996.
[40] [DETAIL_TODO: authors], "Uncertainty-aware state of charge estimation for lithium-ion batteries with gated recurrent unit and deep evidential regression," *IFAC-PapersOnLine*, [DETAIL_TODO: volume, issue, pages, DOI], 2025.
[41] [DETAIL_TODO: authors], "State-of-charge estimation of lithium-ion batteries using ultrasonic-based uncertainty-aware learning models," in [DETAIL_TODO: proceedings title], Springer, 2026, doi: 10.1007/978-981-95-8232-7_61.
[42] J. Xue, X. Du, L. Zhao, Z. Yang, C. Xia, Y. Ma, M. J. Hoque, W. Fu, X. Yan, and N. Miljkovic, "Active-learning-driven error control for data-driven state of charge estimation across the lithium battery lifecycle," *Energy AI*, vol. 21, Art. no. 100549, 2025, doi: 10.1016/j.egyai.2025.100549.
[43] P. Kollmeyer, "Panasonic 18650PF Li-ion battery data," Mendeley Data, v1, 2018, doi: 10.17632/wykht8y7tg.1.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)