Review of Computer Engineering Research
2023 Vol. 10, No. 2, pp. 28-39
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v10i2.3413

An imperative role of studying existing battery datasets and algorithms for battery management system

Gopal Krishna[1]
Rajesh Singh[1,2+]
Anita Gehlot[1,2]
Pardeep Singh[3]
Sandeep Rana[4]
Shaik Vaseem Akram[1]
Kapil Joshi[1]

[1]*Uttaranchal Institute of Technology, Uttaranchal University, Dehradun-248007, Uttarakhand, India.*
[1]*Email: cse.gopalkrishna@gmail.com*
[1]*Email: vaseemakram5491@gmail.com*
[1]*Email: kapilengg0509@gmail.com*
[2]*Department of Project Management, Universidad Internacional Iberoamericana, Campeche 24560, CP, Mexico.*
[1,2]*Email: drrajeshsingh004@gmail.com*
[1,2]*Email: dranitagehlot@gmail.com*
[3]*Graphic Era Hill University, Dehradun-248002, Uttarakhand, India.*
[3]*Email: pardeep.maan@gmail.com*
[4]*G B Pant Institute of Engineering and Technology, Pauri Garhwal-246194 Uttarakhand, India.*
[4]*Email: srana231@gmail.com*

(+ Corresponding author)

Article History
Received: 13 February 2023
Revised: 10 May 2023
Accepted: 20 June 2023
Published: 3 August 2023

Keywords
AI
Battery dataset
BMS
Lithium-ion
ML
SoH.

ABSTRACT

Numerous portable technologies, including electric vehicles, cell phones, and laptops, are powered by batteries. The use of batteries is increasing due to the widespread usage of battery energy storage in the generation of renewable energy. This has also resulted in an increase in the number of negative incidents related to batteries and had a significant negative economic impact on industries. The shortcomings in the traditional monitoring of lithium-ion batteries have been overcome through the implementation of advanced technologies. However, few studies have discussed the use of distinct datasets in the implementation of an intelligent battery management system (BMS). This paper presents a discussion on the choice of dataset variables and applied algorithms for the implementation of effective BMS with artificial intelligence (AI) and machine learning (ML). The study analyzed the use of different datasets, including the National Aeronautics and Space Administration (NASA) battery dataset, to improve BMS. It found that the dataset variables must include the terminal voltage, terminal current, charge current, charge voltage, internal resistance, temperature, and cycle to calculate the state of health (SoH). In future, BMS hardware will be developed to obtain more precise results using AI and ML-based prediction models, utilizing the selected dataset characteristics and variables to achieve longer battery life.

Contribution/Originality: This study investigates the helpfulness of the available battery datasets for training and testing machine learning models and identifies an improved model for predicting a battery's state of health.

1. INTRODUCTION

Renewable utility-scale power generation and storage have become much more cost-effective and important in recent years. Energy storage systems are collections of functions or methods used to collect and store energy [1]. As the demand for energy increases, new energy storage technologies are developed. Storage types include electrochemical, mechanical, chemical, thermal storage, and more. Table 1 shows a comparative analysis of energy storage types with their storage time and efficiency [2].

Table 1. Comparative analysis of energy storage types.

	Storage type	Storage time	Efficiency	Example
1	Electrochemical storage	Short	High	Battery technology
2	Mechanical storage	Long	High	Flywheels
3	Thermal storage	Long	High	Thermochemical energy
4	Chemical storage	Long	Less	H_2 (hydrogen), CH_4 (methane)

Battery technology is the most commonly used form of energy storage as it converts electrochemical energy into electrical energy. Some examples of common battery cells are lithium-ion cells, alkaline cells, carbon-zinc cells, lead acid, zinc-air, and more. Nowadays, lithium-ion batteries dominate the market and research environment because they show better results in the areas of energy efficiency and energy density compared to other batteries. This is the reason for the wide utilization of these batteries in various hand-held devices and electric vehicles [3]. Figure 1 illustrates the various accidents that have occurred due to incorrect data in battery management or inefficient algorithms [4]. The occurrence of these accidents can be minimized by using different algorithms to predict them based on live data. However, predictions based on live data have limitations; they may take a long time and are costly. For this reason, available battery data sets can be used to save time and money. This study evaluated the use of battery datasets and their variables to improve battery management systems (BMS) [5]. Based on the literature, we develop and discuss the vital recommendations that can be applied in future research.

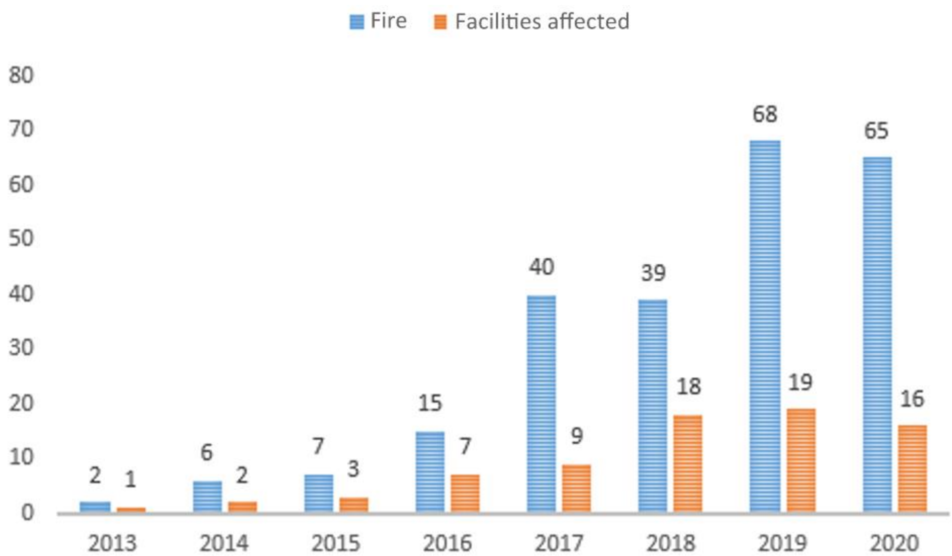

Figure 1. Representation of Li battery fires and facilities affected.

The main contributions of the study are as follows:

• The concepts and types of energy storage are discussed, and a comparative analysis is conducted of the different types of datasets and algorithms based on publicly available battery datasets.

• Based on the analysis, the article discusses the parameters and recommends parameters for future use.

The structure of the paper is as follows. Section 2 discusses the methodology; Section 3 presents an overview of BMS; Section 4 discusses the importance of studying existing battery datasets; Section 5 discusses the available datasets and algorithms; Section 6 covers the analysis of the publicly available datasets and algorithms; Section 7 presents recommendations; Section 8 concludes and suggests future directions.

2. METHODOLOGY

BMS is essential for optimizing the performance and lifespan of batteries in various applications. This study investigated the role of existing battery datasets and BMS algorithms to improve battery management efficiency. The methodology comprised five steps:

- Data collection: The study collected existing battery datasets from academic research papers, battery manufacturers, and open-source repositories. The datasets covered a wide range of battery chemistry and operating conditions relevant to BMS. The collected datasets were used to evaluate existing algorithms and variables and propose improvements to enhance BMS performance.
- Data preprocessing: After collecting the datasets, the study preprocessed them by removing any missing values, outliers, and redundant data points. Normalizing the data ensures consistency across different datasets. The preprocessing step is critical to ensure the accuracy and reliability of the evaluation and analysis of existing algorithms.
- Algorithm and dataset comparison: Based on the evaluation results, the study compared the performance of different algorithms and identified the factors that influence their performance, such as battery chemistry, operating conditions, and system design. The algorithm comparison step provides insight into the best algorithm for specific applications and identifies opportunities for improvement.
- Improvement proposal: Finally, the study proposed required algorithms and variables based on the analysis of the dataset and algorithm performance. The proposed improvements aim to enhance the efficiency and reliability of BMS. Additionally, the study provides recommendations for the design of BMS to improve battery management efficiency.

3. OVERVIEW OF BATTERY MANAGEMENT SYSTEMS

A BMS is a system that is designed to manage and monitor the performance of rechargeable batteries. A BMS helps to ensure that the battery operates safely and efficiently while extending its life. BMS technology has become increasingly important as the demand for rechargeable batteries has grown. Today, BMS technology is used in a wide range of applications, including electric vehicles, renewable energy storage, portable devices, and more [6]. At its core, a BMS is designed to perform several key functions. These functions include monitoring the state of the battery, controlling the charging and discharging process, and protecting the battery from overcharging, overheating, and other potentially harmful conditions [7]. A typical BMS consists of several different components, including a microcontroller or processor, sensors, a battery charger, and other electronic components. These components work together to monitor and control the battery's performance. BMS technology has evolved rapidly in recent years, with new features and capabilities being added all the time. Some of the latest advances in BMS technology include wireless connectivity, advanced analytics, and the ability to integrate with other systems and devices [8].

Each of a BMS' several different components plays a critical role in monitoring and controlling the performance of a rechargeable battery. The following are some of the most important components of a BMS, as illustrated in Figure 2 [9]:

- Microcontroller or processor: The microcontroller or processor is the "brain" of the BMS. It controls the overall operation of the system, including monitoring the state of the battery, controlling the charging and discharging process, and communicating with other components.
- Sensors: Sensors are used to monitor various parameters of the battery, such as voltage, temperature, and current. These sensors provide the data that the BMS needs to determine the state of the battery and make decisions about how to control the charging and discharging process.
- Battery charger: The battery charger is responsible for charging the battery when it is low on charge. The BMS controls the charging process to ensure that the battery is charged safely and efficiently.
- Protection circuits: Protection circuits are used to prevent the battery from becoming damaged due to overcharging, over-discharging, or other potentially harmful conditions. These circuits can shut down the charging process if a problem is detected, preventing further damage to the battery.

- Communication interfaces: Communication interfaces are used to communicate with other systems or devices, such as a vehicle's onboard computer or a renewable energy system. These interfaces allow the BMS to receive instructions or provide data about the battery's performance.
- Balancing circuit: A balancing circuit is used to ensure that the charge is evenly distributed across the individual cells in a battery. This helps to prevent one cell from becoming overcharged or undercharged compared to the others, which can extend the life of the battery.
- Safety features: A BMS may include various safety features such as overvoltage protection, overcurrent protection, and thermal protection to ensure that the battery operates safely and does not pose a risk of damage or injury.

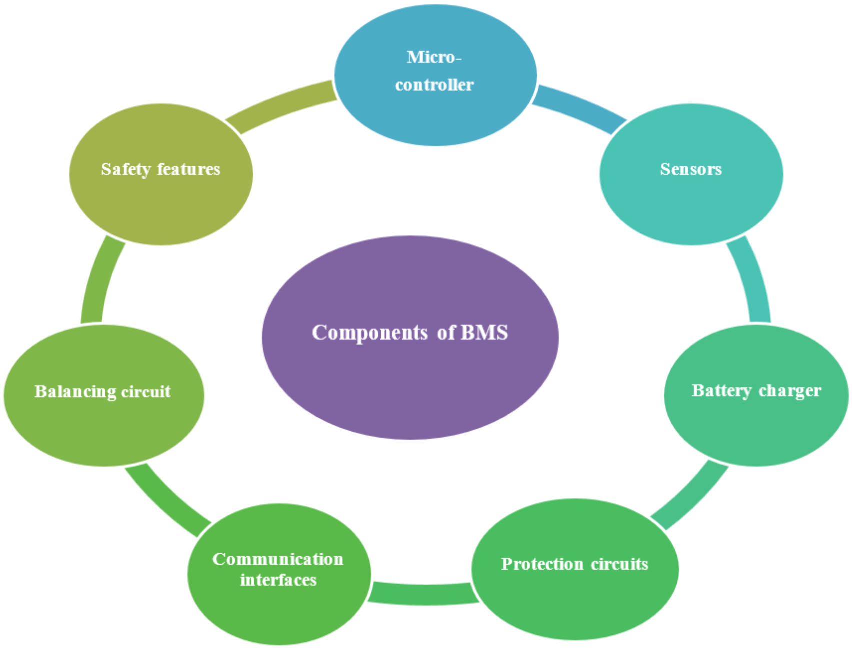

Figure 2. BMS components.

There are several types of BMS, which are used to monitor and control the performance and health of batteries in various applications. Figure 3 illustrates some common types of BMS; these include [7]:

- Passive BMS: A passive BMS is a simple system that uses passive components such as resistors and capacitors to balance the charge across the individual cells in a battery. Passive BMS systems are relatively inexpensive and easy to implement, but they may not be as effective as more advanced systems in balancing the charge.
- Active BMS: An active BMS uses active components such as transistors and amplifiers to balance the charge across the individual cells in a battery. Active BMS systems are generally more effective than passive systems, but they may be more expensive and complex to implement.
- Hybrid BMS: A hybrid BMS combines elements of both passive and active systems, using passive components for some functions and active components for others. Hybrid systems can offer a good balance of effectiveness and cost and are commonly used in many battery applications.
- Distributed BMS: A distributed BMS is a system in which each cell in a battery has its own BMS. This allows for very precise monitoring and control of the battery's performance but can be expensive and complex to implement.
- Centralized BMS: A centralized BMS is a system in which all of the cells in a battery are monitored and controlled by a single BMS. This can be a simpler and less expensive approach than a distributed system but may not offer the same level of precision and control.

- Integrated BMS: An integrated BMS is a system in which the BMS is integrated directly into the battery itself, rather than being a separate component. Integrated BMS systems can be very compact and efficient but may be more difficult to repair or replace if something goes wrong.

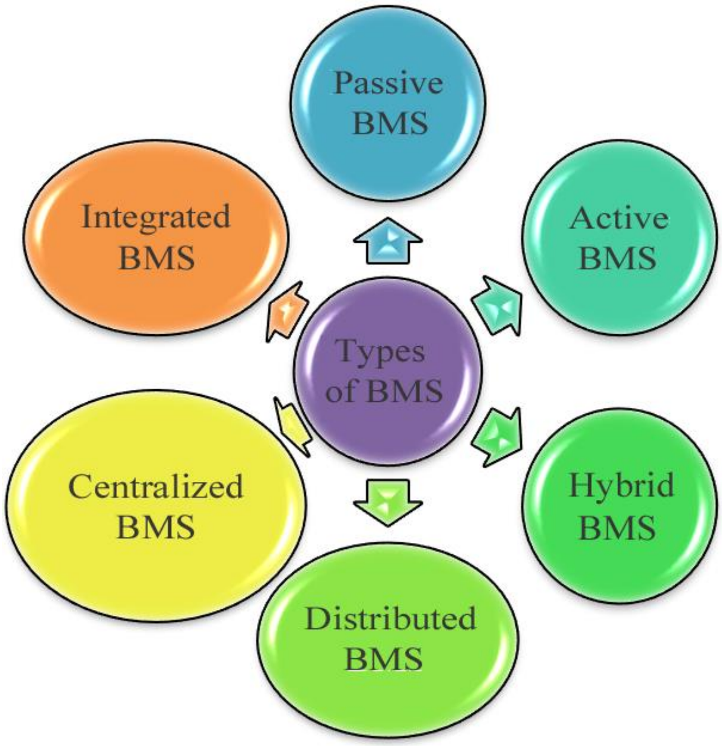

Figure 3. Types of BMS.

Battery management systems are critical components for the proper functioning and safety of rechargeable battery systems. Without a BMS, batteries can become damaged or even dangerous, leading to decreased performance, shorter lifetimes, or even catastrophic failure. One of the key functions of a BMS is to monitor the state of the battery. This includes measuring the battery's voltage, temperature, and other important parameters. By monitoring these parameters, the BMS can detect potential problems early, allowing for preventative action to be taken before more serious issues arise [10].

Another important function of a BMS is to control the charging and discharging process. Proper charging and discharging are critical for the health and longevity of a battery, and a BMS can help to ensure that the battery is charged and discharged safely and efficiently. Additionally, a BMS can protect the battery from overcharging, overheating, and other potentially harmful conditions. Overcharging a battery can lead to thermal runaway, which can cause the battery to catch fire or explode. A BMS can prevent overcharging by monitoring the battery's state and controlling the charging process. In addition to these safety benefits, a BMS can also help to extend the life of a battery. By monitoring and controlling the battery's performance, a BMS can prevent over-discharging or overcharging, which can lead to premature battery failure. Additionally, a BMS can help to balance the charge across the individual cells in a battery, which can prevent one cell from becoming overcharged or undercharged compared to the others [11].

4. IMPORTANCE OF STUDYING EXISTING BATTERY DATASETS

Existing battery datasets provide valuable information on battery performance, lifespan, and reliability under various operating conditions. Analyzing these datasets can help researchers identify the key performance indicators that affect battery behavior, such as temperature, state of charge, and discharge rates. This information can be used to improve battery design, optimize BMS systems, and extend battery lifespan. Studying existing battery datasets can also save time and money compared to conducting new experiments or simulations. Real-world data is often more reliable than simulations, which may not accurately represent the complex interactions that occur within a battery. Access to existing datasets allows researchers to quickly test and validate new ideas and compare their results with those of previous studies. In addition to improving battery performance, studying existing battery datasets can also help researchers identify new applications for batteries. For example, datasets on the energy consumption of buildings or transportation systems can be used to optimize the use of batteries in energy storage and grid stabilization. This can contribute to the development of more sustainable energy systems and reduce the reliance on fossil fuels.

However, there are also challenges associated with analyzing existing battery datasets. Data may be incomplete or inconsistent, making it difficult to draw conclusions or identify patterns. Additionally, datasets may be proprietary or subject to privacy concerns, limiting access to valuable information. Overcoming these challenges requires collaboration between researchers and data providers, as well as the development of standardized data formats and protocols [12].

In short, studying existing battery datasets is essential for advancing battery technology, improving battery performance, and identifying new applications for batteries. While there are challenges associated with analyzing these datasets, the benefits outweigh the costs. Access to real-world data can save time and money and enable researchers to quickly test and validate new ideas. By working together to overcome these challenges, we can continue to make progress toward a more sustainable energy future.

5. AVAILABLE DATASETS

Previous data sets play a very important role in minimizing time and costs, as well as developing new algorithms or comparing existing algorithms. Numerous public data sets are available that can be used for this purpose. Table 2 shows the existing public data sets on batteries.

As Table 2 shows, three datasets use the lithium titanate (LCO) cell chemistry, four datasets use lithium iron phosphate (LFP) cell chemistry, two datasets use nickel cobalt aluminum (NCA) cell chemistry, five datasets use nickel manganese cobalt (NMC) cell chemistry, one dataset uses NMC-LCO cell chemistry, two datasets use NCA, NMC, and LFP cell chemistry, and one dataset uses LCO, LFP, NCA, and NMC cell chemistry. The datasets cover the years 2008 to 2022. The fast-charging dataset uses 230 cells, which is the maximum. The Prognostics Center of Excellence (PCoE) battery dataset was published by NASA.

6. AVAILABLE PARAMETERS OF DATASETS USED IN ALGORITHMS

To calculate the life of a battery, the various data sets employ a few algorithms, such as the fast-charging dataset algorithm, state of health (SoH) estimation algorithm, state of charge (SoC) estimation algorithm, and prognostic algorithm. Figure 4 shows the algorithm used by the different datasets.

In the PCoE battery dataset, the number of cells is 34, the cell form factor is 18650, and the cell chemistry is NCA, as shown in Table 2. This data set was released in 2008-2010. The algorithm used to generate predictions from this dataset is prognostic, and the employed parameters are voltage, current (measured, load), time, and temperature, as shown in Table 3.

In the randomized battery usage dataset, the number of cells is 28, the cell form factor is 18650, and the cell chemistry is LCO, as shown in Table 2. This data set was released in 2014. The algorithm used to generate predictions from this dataset is prognostic, and the employed parameters are voltage, current (measured, load), time, and temperature, as shown in Table 3.

In the Center for Advanced Life Cycle Engineering (CALCE) CS2 dataset, the number of cells is 15, the cell form factor is not provided, and the cell chemistry is LCO, as shown in Table 2. This data set was released in 2010-

2013. The algorithm used to generate predictions from this dataset is SoH estimation, and the employed parameters are time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge), energy (charge, discharge), and internal resistance (AC impedance, ACI phase angle), as shown in Table 3.

Table 2. Data set details.

Ref.	Dataset name	Number of cells	Cell form factor	Cell chemistry	Year
Orzech [13]	Prognostics Center of Excellence (PCoE) battery dataset	34	18650	Nickel cobalt aluminum (NCA)	2008-2010
Bole, et al. [14]	Randomized battery usage dataset	28	18650	Lithium titanate (LCO)	2014
Xing, et al. [15]	Center for Advanced Life Cycle Engineering (CALCE) CS2 dataset	15	-	LCO	2010-2013
Experimental Data Platform [16]	Cycle life prediction dataset	135	18650	Lithium iron phosphate (LFP)	2017-2018
Attia, et al. [17]	Fast-charging optimization dataset	230	18650	LFP	2018-2019
Dubarry and Beck [18]	Synthetic training diagnosis dataset	-	18650 and 26650	LFP	2020
Dubarry and Beck [19]	Short-term cycling performance dataset	-	18650 and 26650	LFP	2020
BatteryArchive [20]	Long-term degradation dataset	86	18650	NCA, nickel manganese cobalt (NMC), and LFP	2018-2020
BatteryArchive [20]	Hawai'i Natural Energy Institute (HNEI) dataset	15	18650	NMC-LCO	2013-2014
Wang, et al. [21]	Oxford battery degradation dataset	8	2018	NMC	2015
Kollmeyer [22]	18650PF dataset (Panasonic)	1	18650	NCA	2018
Automotive Li-ion Cell Usage Data Set [23]	Automotive Li-ion cell usage dataset	1	Pouch cell	NMC	2022
Wang, et al. [21]	Lithium-ion battery (LIB) and Ultracapacitor behavior under dynamic stress test (DST) and urban dynamometer driving schedule (UDDS)	1 pack (containing 4 cells)	Prismatic	LFP	2016
Zhang, et al. [24]	Battery EIS dataset	12	coin cell	LCO	2019
Kollmeyer, et al. [25]	18650PF dataset (Panasonic)	1	18650	NMC	2020
Pozzato, et al. [26]	Aging dataset from electric vehicle (EV) real-driving profiles	10	2170	NMC	2020-2022
Catenaro and Onori [27]	LFP, NMC, NCA battery dataset	18	2170, 1865, 2665	NMC, LFP, NCA	2021
Kollmeyer, et al. [28]	LG 18650HG2	4	18650	NMC	2022

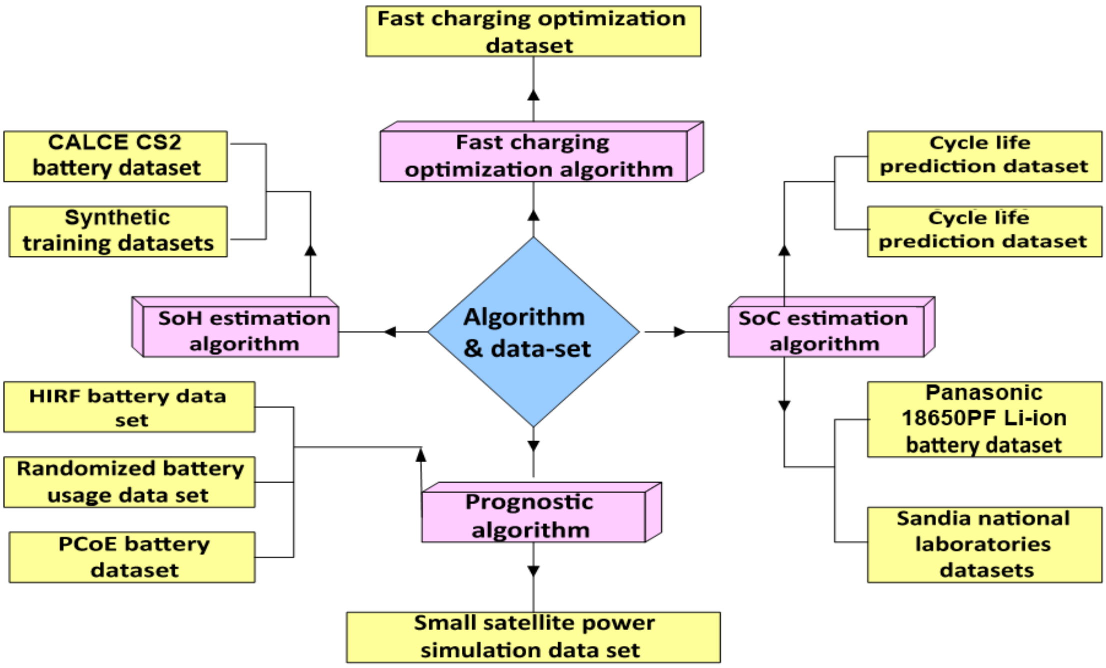

Figure 4. Datasets and algorithms used.

In the cycle life prediction dataset, the number of cells is 135, the cell form factor is 18650, and the cell chemistry is LFP, as shown in Table 2. This data set was released in 2017-2018. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, and temperature, as shown in Table 3. In the fast-charging optimization dataset, the number of cells is 230, the cell form factor is 18650, and the cell chemistry is LFP, as shown in Table 2. This data set was released in 2018-2019. The algorithms used to generate predictions from this dataset are fast-charging optimization and SoC estimation, and the employed parameters are time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, and temperature, as shown in Table 3.

In the synthetic training diagnosis dataset, the number of cells is not given, the cell form factors are 18650 and 26650, and the cell chemistry is LFP, as shown in Table 2. This data set was released in 2020. The algorithm used to generate predictions from this dataset is SoH estimation, and the employed parameters are cycle (charge and discharge), current, voltage, and temperature, as shown in Table 3. In the short-term cycling performance dataset, the number of cells is not given, the cell form factors are 18650 and 26650, and the cell chemistry is LCO, LFP, NCA, and NMC, as shown in Table 2. This data set was released in 2020. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are incremental capacity, normalized capacity, voltage, and degradation, as shown in Table 3. In the long-term degradation dataset, the number of cells is 86, the cell form factor is 18650, and the cell chemistry is NCA, NMC, and LFP, as shown in Table 2. This data set was released in 2018-2020. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are incremental capacity, normalized capacity, voltage, and degradation, as shown in Table 3.

In the Hawai'i Natural Energy Institute (HNEI) dataset, the number of cells is 15, the cell form factor is 18650, and the cell chemistry is NMC-LCO, as shown in Table 2. This data set was released in 2013-2014. The algorithm used to generate predictions from this dataset is a reference performance test, and the employed parameters are temperature (C), max SoC, min SoC, charge rate (C), and discharge rate (C), as shown in Table 3. In the Oxford battery degradation dataset, the number of cells is 8, the cell form factor is 2018, and the cell chemistry is NMC, as shown in Table 2. This data set was released in 2015. The algorithms used to generate predictions from this dataset are drive cycle tests and characterization tests, and the employed parameters are charge-discharge cycle, recorded voltage, current, and temperature, as shown in Table 3. In the 18650PF dataset (Panasonic), the number of cells is 1, the cell form factor is 18650, and the cell chemistry is NCA, as shown in Table 2. This data set was released in 2018. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are time stamp, step, status, prog time, step time, and cycle, as shown in Table 3.

In the automotive Li-ion cell usage dataset, the number of cells is 1, the cell form factor is pouch cell, and the cell chemistry is NMC, as shown in Table 2. This data set was released in 2022. The algorithm used to generate predictions from this dataset is the federal test, and the employed parameters are charge, discharge, and impedance, as shown in Table 3. In the lithium-ion battery and ultracapacitor behavior under dynamic stress test (DST) and urban dynamometer driving schedule (UDDS) dataset, the number of cells is 1 pack (containing 4 cells), the cell form factor is prismatic, and the cell chemistry is LFP, as shown in Table 2. This data set was released in 2016. The algorithm used with this dataset is SoC prediction, and the employed parameters are battery current, battery voltage, ultracapacity current, ultracapacity voltage, and time, as shown in Table 3. In the battery EIS dataset, the number of cells is 12, the cell form factor is coin cell, and the cell chemistry is LCO, as shown in Table 2. This data set was released in 2019. The algorithms used to generate predictions from this dataset are SoH and remaining useful life (RUL) prediction, and the parameters are current, voltage, temperature, and time, as shown in Table 3.

In the 18650PF dataset (Panasonic), the number of cells is 1, the cell form factor is 18650, and the cell chemistry is NMC, as shown in Table 2. This data set was released in 2020. The algorithm used to generate predictions from this dataset is SoC estimation, and the used parameters are time stamp, step, status, prog time, step time, and cycle, as shown in Table 3. In the aging dataset from EV real-driving profiles, the number of cells is 10, the cell form factor is 2170, and the cell chemistry is NMC, as shown in Table 2. This data set was released in 2020-2022. The algorithm used to generate predictions from this dataset is EIS tests, and the employed parameters are time (date, test, step), step index, cycle index, current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, and aux temperature, as shown in Table 3.

In the LFP, NMC, NCA battery dataset., the number of cells is 18, the cell form factors are 2170, 1865, and 2665, and the cell chemistry is NMC, LFP, NCA, as shown in Table 2. This data set was released in 2021. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are current, voltage, surface temp, time (date, test, step), and step index, as shown in Table 3. In the LG 18650HG2 dataset, the number of cells is 4, the cell form factor is 18650, and the cell chemistry is NMC, as shown in Table 2. This data set was released in 2020. The algorithm used to generate predictions from this dataset is SoC estimation, and the employed parameters are step, status, prog time, step time, cycle, cycle level, procedure, voltage, current, temperature, capacity, cells, maximum voltage, gassing voltage, break voltage, charge factor, impedance, cold cranking, and current energy density, as shown in Table 3.

Table 3. Dataset parameters and algorithm used.

Ref.	Dataset name	Algorithm	Data variables/parameters
Gabbar, et al. [6]	PCoE battery dataset	Prognostic	Voltage, current (measured, load), time, temperature
Uzair, et al. [7]	Randomized battery usage dataset	Prognostic	
Ramkumar, et al. [8]	CALCE CS2 dataset	SoH estimation	Time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge, energy (charge, discharge), internal resistance, AC impedance, ACI phase angle
Lipu, et al. [9]	Cycle life prediction dataset	SoC estimation	Time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, temperature
Uzair, et al. [7]	Fast-charging optimization dataset	Fast-charging optimization, SoC estimation	

Ref.	Dataset name	Algorithm	Data variables/parameters
See, et al. [10]	Synthetic training diagnosis dataset	SoH estimation	Cycle (charge and discharge), current, voltage, temperature
Habib, et al. [11]	Short-term cycling performance dataset	SoC estimation	Incremental capacity, normalized capacity, voltage, degradation
Dos Reis, et al. [12]	Long-term degradation dataset	SoC estimation	Incremental capacity, normalized capacity, voltage, degradation
Dos Reis, et al. [12]	HNEI dataset	Reference performance test	Temperature (C), max SoC, min SoC, charge rate (C), discharge rate (c)
Orzech [13]	Oxford battery degradation dataset	Drive cycle tests and characterization tests	Charge-discharge cycle, recorded voltage, current, temperature
Bole, et al. [14]	18650PF dataset (Panasonic)	SoC estimation	Time stamp, step, status, prog time, step time, cycle
Xing, et al. [15]	Automotive li-ion cell usage dataset	Federal test	Charge, discharge, impedance
Experimental Data Platform [16]	LIB and ultracapacitor behavior under DTS and UDDS	SoC prediction	Battery current, battery voltage, ultra capacity current, ultra capacity voltage, time
Attia, et al. [17]	Battery electrochemical impedance spectroscopy (EIS) dataset	SoH, RUL prediction	Current charge, voltage charge, temperature, time
Dubarry and Beck [18]	18650PF dataset (Panasonic)	SoC estimation	Time stamp, step, status, prog time, step time, cycle
Dubarry and Beck [19]	Aging dataset from EV real-driving profiles	EIS tests	Time (date, test, step), step index, cycle index, current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, aux temperature
BatteryArchive [20]	LFP, NMC, NCA battery dataset	SoC estimation	Current, voltage, surface temp, time (date, test, step), step index
Wang, et al. [21]	LG 18650HG2	SoC estimation	Step, status, prog time, step time, cycle, cycle level, procedure, voltage, current, temperature, capacity, cells, maximum voltage, gassing voltage, break voltage, charge factor, impedance, cold cranking, current energy density

7. RECOMMENDATIONS

In the previous section, we detailed the various battery dataset options available for battery-management systems to choose from to make correct predictions of battery performance. Based on the study findings, we make the following recommendations:

1. For a BMS to obtain the battery statistics, the parameters must include time (test, date, step), index (step, cycle), current, voltage, capacity (charge, discharge), energy (charge, discharge), internal resistance, and temperature.
2. To calculate a battery cell's state of health (SoH) or aging, the algorithm must include the calculation of SoC and depth of discharge (DoD).

8. CONCLUSIONS AND FUTURE DIRECTIONS

Due to the widespread use of renewable energy sources and the desire for sustainability, battery-management systems have attracted a lot of attention. Battery fitness monitoring is essential for reliably storing power. Approaches to estimating battery fitness have been developed for monitoring the final ability and electricity estimation, ability prediction, lifestyle and fitness prediction, as well as critical indications related to battery stability and thermal management. From the review of the numerous available datasets, we have concluded which parameters should be selected for the collection of data from batteries. Depending on the algorithm used, the datasets include different attributes; however, essential variables such as current, voltage, temperature, and time stamp remain the same. Additionally, we recommend including certain parameters when building new datasets standards for the algorithms, especially for battery safety. Finally, this paper suggests recommendations for battery parameters that can be collected through BMS. By feeding accurate statistics to the ML-based algorithm, accidents can be minimized by reducing overcharging, deep discharging, and overheating of the battery, as well as predicting the age of the battery and increasing its life span.

Funding: This research is supported by Uttaranchal University, Uttarakhand, India under the SEED money project 2021–2022 (Grant number: UU/DRI/SM/2021/05).
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Conceptualization, G.K. and R.S.; methodology, A.G.; formal analysis, S.V.A.; investigation, S.R.; resources, P.S.; data curation, R.S. and S.R.; writing, R.S. and S.V.A.; writing, review and editing, A.G.; visualization, P.S; supervision, R.S.; funding acquisition, K.J. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] G. Krishna, R. Singh, A. Gehlot, S. V. Akram, N. Priyadarshi, and B. Twala, "Digital technology implementation in battery-management systems for sustainable energy storage: Review, challenges, and recommendations," *Electronics*, vol. 11, no. 17, p. 2695, 2022. https://doi.org/10.3390/electronics11172695

[2] S. A. Hasib *et al.*, "A comprehensive review of available battery datasets, RUL prediction approaches, and advanced battery management," *Ieee Access*, vol. 9, pp. 86166-86193, 2021. https://doi.org/10.1109/access.2021.3089032

[3] Lithium-Ion Battery, "Lithium-Ion battery - clean energy institute," Retrieved: https://www.cei.washington.edu/education/science-of-solar/battery-technology/. [Accessed Jan. 27, 2023], 2020.

[4] The Importance of Sending Consumers, "The importance of sending consumers' used lithium-ion batteries to electronic recyclers or hazardous waste collection facilities | US EPA," Retrieved: https://www.epa.gov/recycle/importance-sending-consumers-used-lithium-ion-batteries-electronic-recyclers-or-hazardous. [Accessed Jan. 27, 2023], 2021.

[5] S. R. Niraula, "A review of research process, data collection and analysis," *Insights in Biology and Medicine*, vol. 3, no. 1, pp. 001-006, 2019. https://doi.org/10.29328/journal.ibm.1001014

[6] H. A. Gabbar, A. M. Othman, and M. R. Abdussami, "Review of battery management systems (BMS) development and industrial standards," *Technologies*, vol. 9, no. 2, p. 28, 2021. https://doi.org/10.3390/technologies9020028

[7] M. Uzair, G. Abbas, and S. Hosain, "Characteristics of battery management systems of electric vehicles with consideration of the active and passive cell balancing process," *World Electric Vehicle Journal*, vol. 12, no. 3, p. 120, 2021. https://doi.org/10.3390/wevj12030120

[8] M. S. Ramkumar *et al.*, "Review on Li-Ion battery with battery management system in electrical vehicle," *Advances in Materials Science and Engineering*, vol. 2022, p. 8, 2022. https://doi.org/10.1155/2022/3379574

[9] M. S. H. Lipu *et al.*, "Battery management, key technologies, methods, issues, and future trends of electric vehicles: A pathway toward achieving sustainable development goals," *Batteries*, vol. 8, no. 9, p. 119, 2022. https://doi.org/10.3390/batteries8090119

[10] K. See *et al.*, "Critical review and functional safety of a battery management system for large-scale lithium-ion battery pack technologies," *International Journal of Coal Science & Technology*, vol. 9, no. 1, p. 36, 2022. https://doi.org/10.1007/s40789-022-00494-0

[11] A. A. Habib, M. K. Hasan, G. F. Issa, D. Singh, S. Islam, and T. M. Ghazal, "Lithium-ion battery management system for electric vehicles: Constraints, challenges, and recommendations," *Batteries*, vol. 9, no. 3, p. 152, 2023. https://doi.org/10.3390/batteries9030152

[12] G. Dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy and AI*, vol. 5, p. 100081, 2021. https://doi.org/10.1016/j.egyai.2021.100081

[13] G. Orzech, "Prognostics center of excellence data set repository," Retrieved: http://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository. [Accessed Jan. 27, 2023], 2022.

[14] B. Bole, C. S. Kulkarni, and M. Daigle, "Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use," presented at the In Annual Conference of the PHM Society, 2014.

[15] Y. Xing, W. He, M. Pecht, and K. L. Tsui, "State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures," *Applied Energy*, vol. 113, pp. 106-115, 2014. https://doi.org/10.1016/j.apenergy.2013.07.008

[16] Experimental Data Platform, "Experimental data platform," Retrieved: https://data.matr.io/1/projects/5c48dd2bc625d700019f3204. [Accessed Jan. 27, 2023], 2019.

[17] P. M. Attia *et al.*, "Closed-loop optimization of fast-charging protocols for batteries with machine learning," *Nature*, vol. 578, no. 7795, pp. 397-402, 2020. https://doi.org/10.1038/s41586-020-1994-5

[18] M. Dubarry and D. Beck, "Analysis of synthetic voltage vs. capacity datasets for big data Li-ion diagnosis and prognosis," *Energies*, vol. 14, no. 9, p. 2371, 2021. https://doi.org/10.3390/en14092371

[19] M. Dubarry and D. Beck, "Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis," *Journal of Power Sources*, vol. 479, p. 228806, 2020. https://doi.org/10.1016/j.jpowsour.2020.228806

[20] BatteryArchive, "BatteryArchive.org," Retrieved: https://www.batteryarchive.org/snl_study.html. [Accessed Jan. 27, 2023]. 2023.

[21] Y. Wang, C. Liu, R. Pan, and Z. Chen, "Experimental data of lithium-ion battery and ultracapacitor under DST and UDDS profiles at room temperature," *Data in Brief*, vol. 12, pp. 161-163, 2017. https://doi.org/10.1016/j.dib.2017.01.019

[22] P. Kollmeyer, "Panasonic 18650pf li-ion battery data," *Mendeley Data*, vol. 1, no. 2018, 2018.

[23] Automotive Li-ion Cell Usage Data Set, "Automotive Li-ion cell usage data set | IEEE DataPort," Retrieved: https://ieee-dataport.org/documents/automotive-li-ion-cell-usage-data-set. [Accessed Jan. 27, 2023], 2022.

[24] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," *Nature Communications*, vol. 11, no. 1, p. 1706, 2020. https://doi.org/10.1038/s41467-020-15235-7

[25] P. Kollmeyer, M. Naguib, and M. Skells, "LG 18650HG2 Li-ion battery data - mendeley data," Retrieved: https://data.mendeley.com/datasets/b5mj79w5w9/2. [Accessed Jan. 28, 2023], 2020.

[26] G. Pozzato, A. Allam, and S. Onori, "Lithium-ion battery aging dataset based on electric vehicle real-driving profiles," *Data in Brief*, vol. 41, p. 107995, 2022. https://doi.org/10.1016/j.dib.2022.107995

[27] E. Catenaro and S. Onori, "Experimental data of lithium-ion batteries under galvanostatic discharge tests at different rates and temperatures of operation," *Data in Brief*, vol. 35, p. 106894, 2021. https://doi.org/10.1016/j.dib.2021.106894

[28] P. Kollmeyer, C. Vidal, M. Naguib, and M. Skells, "Lg 18650HG2 li-ion battery data and example deep neural network xEV SOC estimator script," *Mendeley Data*, vol. 3, p. 2020, 2020. https://doi.org/10.17632/cp3473x7xv.2