Chapter 1
# Introduction to Battery Full-Lifespan Management

As one of the most promising alternatives to effectively bypass fossil fuels and promote net-zero carbon emission target around the world, rechargeable lithium-ion (Li-ion) batteries have become a mainstream energy storage technology in numerous important applications such as electric vehicles, renewable energy storage, and smart grid. However, Li-ion batteries present inevitable ageing and performance degradation with time. To ensure efficiency, safety and avoid potential failures for Li-ion batteries, reliable battery management during its full-lifespan is of significant importance. This chapter first introduces the background and motivation of Li-ion battery, followed by the description of Li-ion battery fundamentals and the demands of battery management. After that, the basic information and benefits of using data science technologies to achieve effective battery full-lifespan management are presented.

1.1 Background and Motivation

1.1.1 Energy Storage Market

According to the statistics from the CNESA Global Energy Storage Projects Database, the global operating energy storage project capacity has reached 191.1GW at the end of 2020, a year-on-year increase of 3.4% [1]. As illustrated in Fig. 1.1, pumped storage contributes to the largest portion of global capacity with 172.5GW, a year-on-year increase of 0.9%. Electrochemical energy storage becomes the second-largest portion with a total capacity of 14.1GW. Among different electrochemical energy storage solutions, Li-ion batteries reach the capacity of 13.1GW, exceeding 10GW for the first time.
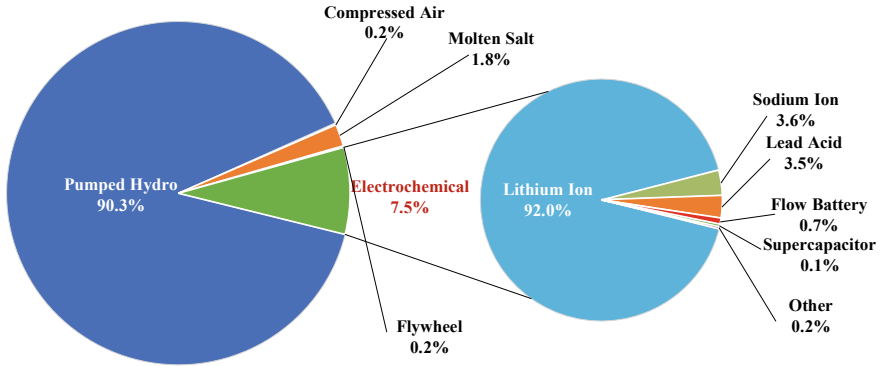
Fig. 1.1 2020 global energy storage market classification share, reprinted from [1], open access

China, the USA, Europe, and Australia are the leaders of the energy storage market. The new operating capacity of these countries accounted for over 86%, which has exceeded the GW-level new operating capacity.

China: Driven by the Chinese policies to encourage and require storage allocation in the energy, the largest installed capacity of new energy power generation in China exceeds 580 MW, a rapid increase of 438%. Furthermore, the establishment of "Carbon Peak" and "Carbon Neutral" targets in China also significantly boosts the leapfrog development of renewable energy and related battery-based energy storage.

In April 2021, the National Development & Reform Commission and the National Energy Administration issued the "Guiding Opinions on Accelerating the Development of New Energy Storage (Draft for Comment)" [2]. For the first time, the document clarifies the development goals of the energy storage industry. By 2025, the installed capacity of new energy storage capacity will reach more than 30 million kilowatts (30 GWh). As of 2020, the cumulative installed capacity of new electric energy storage in China has reached 3.28Gwh [3], a year-on-year increase of 91.2%, which also means that by 2025, the scale of the Chinese new energy storage market will be about 10 times larger than the level at the end of 2020.

USA: A breakthrough has been made in the deployment before the schedule in 2020, and the newly added operating capacity in the USA has doubled in comparison with that in 2019. The newly installed capacity is mainly concentrated in California, while LS Power and Vistra Energy added 250 MW/250MWh and 300 MW/1200MWh projects, respectively. The latter is the largest battery-based energy storage project in the USA and even the world. Besides, the deployment of large-scale 100 MW battery-based energy storage projects in Texas, New York, Florida, and other states has been accelerated.

The National Renewable Energy Lab uses the Regional Energy Deployment System (ReEDS) capacity expansion model to understand the complex dynamics involved with the future market potential for utility-scale energy storage in the contiguous USA [4]. The battery storage mandates enacted in Oregon, California,
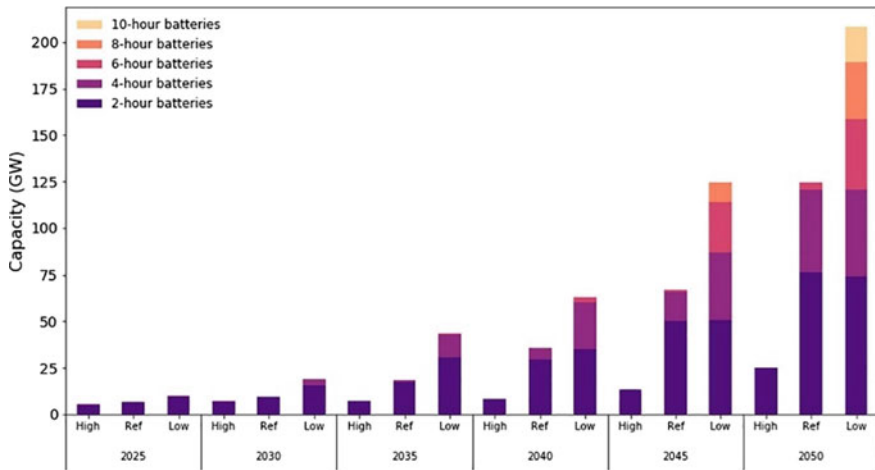
Fig. 1.2 Cumulative model-deployed battery storage with High, Reference, and Low battery capital costs. All scenarios show here use the dynamic assessment of storage capacity credit, reprinted from [4], with permission from Elsevier

New Jersey, New York, and Massachusetts are included. In total, these mandates require the model to build 1775 MW of batteries by 2020, 4685 MW by 2025, and 6555 MW by 2030. As illustrated in Fig. 1.2, the results of the "Reference" and "Low" battery cost scenarios generated by the ReEDS show a significant new battery storage deployment, with the deployment levels of 125 GW and 208 GW in 2050, respectively.

Europe: The implementation of the "Clean Energy for All" program has sent a significantly positive signal for the European energy storage market. This is reflected in the strong performance of the front-end energy storage market for electricity meters in the UK and the strong performance of the home energy storage market in Germany. The UK has cancelled project capacity restrictions, allowed more than 50 and 350 MW projects in England and Wales, and officially launched the construction of large-scale energy storage projects in the UK. With more than 300,000 household battery systems installed in Germany, COVID-19 has further stimulated consumer demand for energy flexibility, safety, and independence.

The European Union (EU) energy and climate policy aims to significantly cut CO_2 emissions in the power sector by 2030 and to establish a nearly carbon-free electricity sector by 2050. The role of transmission and energy storage in European decarburization towards 2050 provides support to the hypothesis that the EU energy and climate targets for 2050 will increase the capacity of intermittent power, storage technologies and international transmission lines, as illustrated in Fig. 1.3. In 2050, the investment in electric battery capacity will range from 80 to 351 GWh [5].

An assessment of European electricity arbitrage using storage systems shows that, in the near future, the most attractive European countries for the electricity
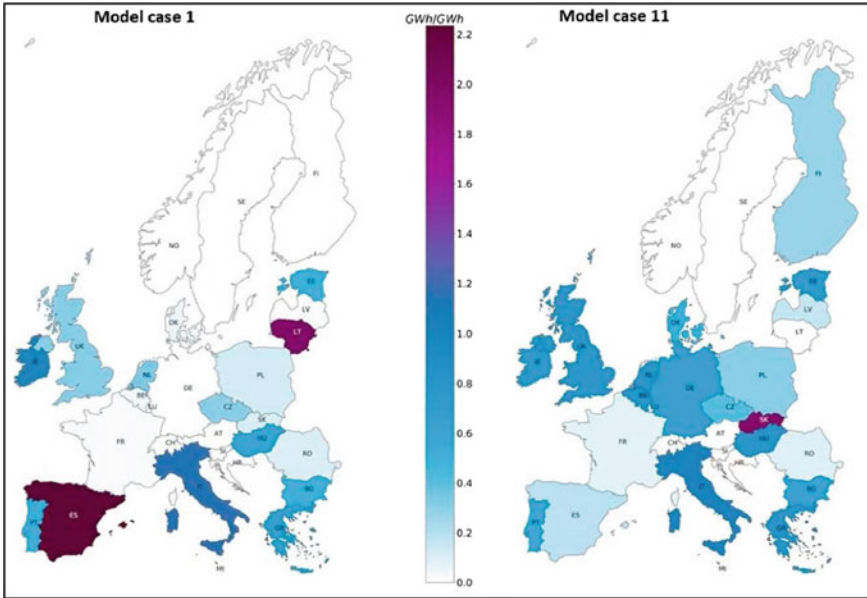
Fig. 1.3 Battery energy capacity per average hourly electricity demand in 2050 for the European countries, reprinted from [5], open access

arbitrage business are the UK and Ireland, with current Net Present Value close to −400,000 V, while Spain and Portugal might show the worst performances, their current Net Present Value are close to −800,000 V.

United Kingdom: The UK was the country with the largest new operating energy storage capacity in the European market, accounting for 44.6% of the total European continent in 2019. In 2019, the UK government signed a legally binding commitment to bring all greenhouse gas emissions to net-zero by 2050. Batteries will play a significant part in this transition, both in transport and renewable energy storage. Improving understanding of how batteries will age and how to design more efficient battery management system will aid the net-zero transition and reduce waste. Batteries and electric vehicles also represent part of the government's industrial strategy for the future of mobility and the mission to "Put UK at the forefront of the design and manufacturing of zero-emission vehicles" [6]. The UK automotive industry adds £18.6 billion to the UK economy, employees 823,000 people, and accounts for 14.4% of all UK goods exports [7]. To facilitate the transition to electrification, £1 billion is being invested in the Advanced Propulsion Centre (APC) and £246 million in the Faraday battery challenge.

Australian: The Australian Energy Market Operator (AEMO) reports that there are 85 big batteries with a total capacity of 18,660 MW in the planning pipeline [8]. How many of these projects can be realized will be a function of battery cost development, as well as the development of different revenue streams that batteries are enabled to

provide. In Australia, batteries can provide revenue-generating services in various markets, most of which are ancillary and wholesale markets [9]. Virtual transmission lines, or avoided transmission investment, are emerging as a potential income stream for batteries.

1.1.2   Li-Ion Battery Role

Demand for Li-ion batteries to power electric vehicles and energy storage has seen exponential growth, increasing from just 0.5 gigawatt-hours in 2010 to around 526 gigawatt hours a decade later. Demand is projected to increase 17-fold by 2030.

According to the commissioned manufacturing capacity of Li-ion battery by plant location in Fig. 1.4, Asia dominates the Li-ion battery supply chain, especially China [10], where Chinese Li-ion battery manufacturer CATL is one of the world leaders in battery manufacturing in 2020, as illustrated in Fig. 1.5. China's success results from its sizeable domestic battery demand, control of more than 70% of the world's graphite raw material refining, and massive cell and cell component manufacturing capacity. Korea and Japan rank number two and three in the Li-ion battery supply chain. While both countries are among the leaders in battery and cell component manufacturing (LG Energy Solution, Samsung SDI, SK Innovation, Panasonic), they do not have the same influence in raw materials refining and mining as China.

Figure 1.6 illustrates the Li-ion battery cell manufacturing capacity by country or region. Obviously, China is the largest battery manufacturing country with 567 GWh, which is nearly ten times larger than the second one-the United States. Europe owns
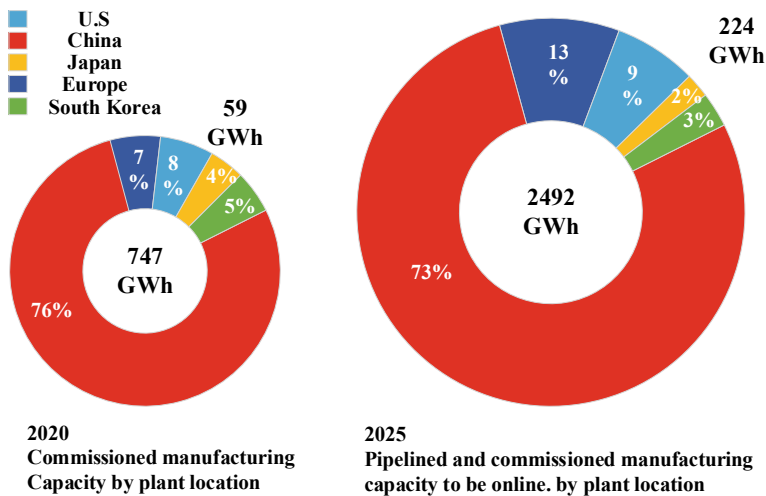

Fig. 1.4 Commissioned manufacturing capacity of Li-ion battery by plant location, 2020 and 2025, reprinted from [10], open access
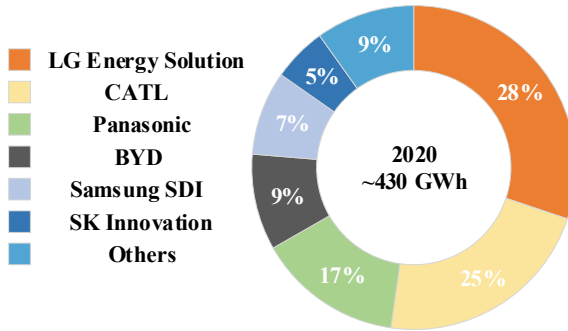
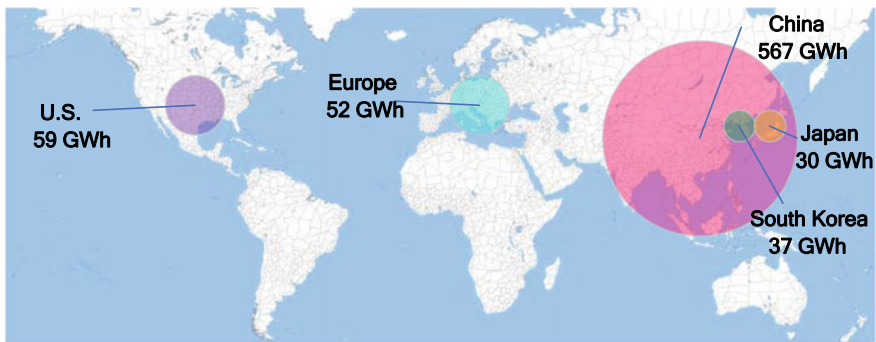**Fig. 1.5** 2020 top battery manufacturers market shares in GWh

Fig. 1.6 Cell manufacturing capacity by country or region, reprinted from [10], open access

the third-largest battery manufacturing capacity. Apart from China, two Asian countries including South Korea and Japan present the fourth and fifth Li-ion battery manufacturing capacity with 37 GWh and 30 GWh, respectively. The detailed percentages of total manufacturing capacity for various battery components by country are listed in Table 1.1.

As analysed by Yole's team in the new Status of the Rechargeable Li-ion Battery Industry 2021 report, Li-ion battery has become the technology of choice for many applications. As a result, it attracts numerous players: R&D labs, cell component manufacturers, cell and battery pack manufacturers, and system integrators. Li-ion battery market is composed of multiple applications of battery technology, with slightly different targets and roles, further resulting in each application being best served by the specific Li-ion battery technology. Three real applications including electric vehicles, electronic devices, and stationary battery-based energy storage comprise the bulk of the current Li-ion battery market, as shown in Fig. 1.7.

Different electric vehicles (xEVs): The rapidly growing xEVs market consists of different types of EVs such as hybrid electric vehicle (HEV), plug-in hybrid electric vehicle (PHEV), full electric vehicle (EV), and commercial electric vehicle (CEV),

Table 1.1 Percentage of total manufacturing capacity of different countries for various components of Li-ion battery (data from [10], open access)

Country	Cathodes manufacturing (3 M tons) (%)	Anode manufacturing (1.2 M tons) (%)	Electrolyte solution manufacturing (339,000 tons) (%)	Separator manufacturing (1987 M m^2) (%)
United States	–	10	2	6
China	42	65	65	43
Japan	33	19	12	21
Korea	15	6	4	28
Rest of World	10	–	17	2

Fig. 1.7 Main applications of Li-ion battery: **a** different electric vehicles, **b** electronic device, **c** stationary battery-based energy storage

as illustrated in Fig. 1.7a, where Li-ion battery plays specific roles in different applications. For HEV, as it belongs to the traditional internal combustion engine-based vehicle, its propulsion system is combined with a small electric motor driven by batteries and these batteries are usually charged through regenerative braking. In this context, the capacity of the Li-ion battery is relatively smaller, further making its energy density and capital cost become less relevant. However, due to frequent braking of HEV, battery here requires to be charged and discharged powerfully. Therefore, Li-ion battery needs to have high power density, quick charging speed and long lifetime over thousands of cycles in HEV. In comparison with HEV, batteries within PHEV could be also charged through plugging into an external electricity source. The battery here generally presents much larger capacity to enable PHEV to drive fully electric for a short distance. In this context, Li-ion battery requires better energy density and lower capital cost, while its power density as well as lifetime

become of less concern. For a full EV without any internal combustion engine, in order to deliver enough ranges for drivers, Li-ion battery generally needs low capital cost and high energy density. Besides, as EV could not fall back on the internal combustion engine anymore, Li-ion battery also needs to present high reliability and long service life over 1000 cycles. For CEV such as e-bus that battery systems are relatively larger and the effects of battery fault such as thermal runaway would become more severe, Li-ion battery here has increased safety requirements. Besides, as e-buses generally need to be charged frequently, the service life of Li-ion battery here also becomes more important than other xEV cases.

Electronic device: Li-ion battery is also widely used to support power/energy for electronic devices such as cell phones, laptops, and tablets, as shown in Fig. 1.7b. Li-ion battery presents the similar roles in all these electronic device applications to provide as much energy as possible in a compact form, so the volumetric energy density becomes the most crucial element. Besides, the cost of Li-ion battery in electronic devices is relatively smaller and users are generally willing to pay for high-performance Li-ion battery, cost here thus becomes the secondary important element. Furthermore, as electronic device application usually presents low drain, the power density of Li-ion battery here becomes less of concern.

Stationary battery-based energy storage (BES): BES is becoming a vital part to smooth the supply and demand of power generated from renewable energy such as wind sources and solar sources, as illustrated in Fig. 1.7c. In real applications, BES ensures the electricity transferred from renewable energy could be stored for further reutilization. Besides, it is able to also ensure that the peak in consumption is absorbed while backup could be provided without having to temporarily rely on fossil fuel power plants, further bringing positive environmental and economic impacts. There are different types of operating models for Li-ion batteries in BES applications. Based upon the requirements of Li-ion battery, these operating models could be divided across two axes as the frequency of discharge and the length of discharge, where the applications and key needs of Li-ion battery in four related quadrants of BES are illustrated in Fig. 1.8.

1.2 Li-Ion Battery and Its Management

1.2.1 Li-Ion Battery

Li-ion battery belongs to an electrochemical energy storage system, which generates a potential difference and allows current to flow through the circuit until the energy is exhausted. The first Li-ion battery was commercially introduced by Sony company in 1991. As illustrated in Fig. 1.9, Li-ion battery consists of three components including anode (negative electrode), cathode (positive electrode), and electrolyte. The active material is bonded to the metal fluid at both ends of the cell and electrically isolated

		Regulation/ Stabilization of: - Frequency -Voltage -Renewables	Peak shaving Load shifting Self-supply & Tou Congestion avoidance
Frequency of discharge	High	Low cost per cycle High power Fast charging	High energy Low cost per cycle
	Low	Frequency response Uninterrupted power supply Black start	Uninterrupted power supply Black start
		Low capital cost High power	Low capital cost High energy
Application		Low	High
Key needs		Length of discharge	

Fig. 1.8 Applications and key needs of Li-ion battery in four related quadrants of stationary battery-based energy storage
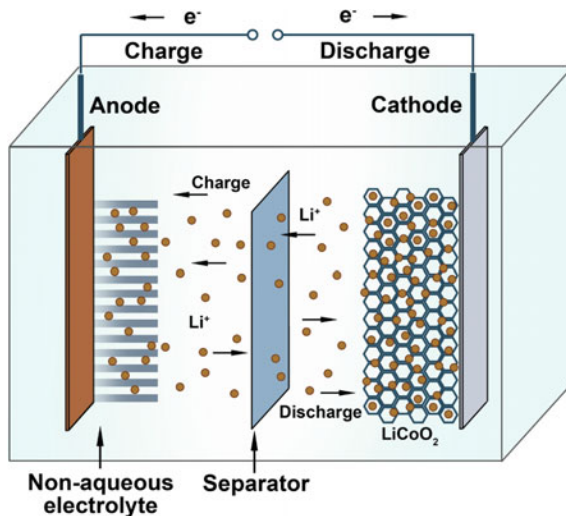

**Fig. 1.9** Operating scheme of Li-ion battery

with a microporous polymer separator or gel polymer. Liquid or gel polymer electrolytes allow lithium ions (Li^+) to diffuse between the positive and negative electrodes. Li-ions are intercalated or deintercalated from the active material through an intercalation process.

The anode here mainly contains graphite. Besides, Li-Titanate anode combined with any other cathode is also developed to provide better safety and battery performance at the sacrifice of energy density. For cathode, it mainly consists of a metal oxide. Among different types of the cathode, lithium cobalt oxide (LCO) is able to offer higher energy density but presents a higher safety risk level, especially when it is damaged. In this context, this chemical composition has been widely adopted in consumer electronics. In contrast, lithium iron phosphate (LFP), lithium manganese oxide (LMO), and lithium nickel manganese cobalt oxide (NMC) batteries would offer lower energy densities, but become inherently safer. The electrolyte mainly consists of a lithium salt in an organic solvent.

Table 1.2 illustrates and summarizes various chemical compositions that are adopted for battery cathode electrode. It can be noted that for Li-ion battery with different materials, its performance such as the voltage, energy density, service life, and safety level could become significantly different. Most metal oxide electrodes are thermally unstable and could decompose at high-temperature conditions, further

Table 1.2 Rechargeable Li-ion batteries with various cathode compositions

Name	LCO	LMO	NMC	NCA	LFP
Full name	Lithium cobalt oxide	Lithium manganese oxide	Lithium nickel manganese cobalt oxide	Lithium nickel cobalt aluminium oxide	Lithium iron phosphate
Cathode	$LiCoO_2$	$LiMn_2O_4$	$LiNiMnCoO_2$	$LiNiCoAlO_2$	$LiFePO_4$
Cell voltage [V]	3.7	3.6	3.8	3.6	3.3
Cut-off voltage [V]	4.2	4.2	4.1–4.3	4.3–4.5	3.6
Energy density [Wh/kg]	150–250	100–170	150–220	200–250	80–140
Thermal runaway temperature [°C]	150	250	200	175	250
Common applications	Cellphone, camera, laptop	Hand tool, medical equipment	Hand tool, medical equipment	Vehicle	Hand tool, medical equipment
Comments	High energy density	Low internal impedance	High energy density Withstands rapid charge	High energy density High cell voltage	Long charge lifetime Withstands rapid charge

| Prismatic | Pouch | Cylindrical |
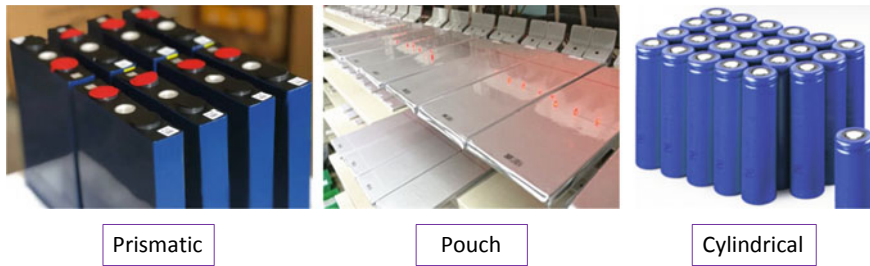
Fig. 1.10 Li-ion cell designs with various shapes

releasing oxygen to result in thermal runaway conditions. Among all these electrode chemical compositions, lithium manganese oxide (LMO) and lithium nickel manganese cobalt oxide (NMC) become the best candidates to compromise between performance and safety levels currently available on the Li-ion battery market.

Besides, Li-ion battery could be designed with different shapes including prismatic, pouch, and cylindrical, as shown in Fig. 1.10. Among these three designs, the prismatic battery cell design becomes the safest one because it is equipped with some mechanisms, such as safety functional layers, multi-layer partitions, safety vents, safety fuses, and overcharge safety devices.

1.2.2 Demands for Battery Management

(1) Battery management system market

Due to the dramatically increased requirements of battery being used in numerous applications such as transportation electrifications and smart grid energy storage, the global market of battery management system also grows rapidly with a compound annual growth rate of over 10%. Here the transportation sector is leading the main market growth of battery management system as a great number of EVs being manufactured and sold annually. For example, the global EV fleet stock was around 10.2 million in 2020, an increase of over 43% in comparison with that in 2019. Battery management plays a pivotal role in determining battery efficiency, performance and safety, especially for EV applications. Therefore, battery management system must be well equipped in an EV.

Figure 1.11 illustrates the growth rate of battery management system market around the world from 2021 to 2026 estimated by the Mordor Intelligence [11]. It can be seen that Asia–Pacific owns the biggest market share for battery management system, mainly due to the dramatically rising sale of EVs in countries like China and Japan. This dramatic increase of EVs is mainly caused by the extensive efforts of the governments to decrease greenhouse gas emissions. For example, China has become the biggest EV market around the world. The market share of Chinese EVs has risen from about 23% in 2015 to 44% in 2020, and by the end of 2020, about 4.5 million

Fig. 1.11 Estimated growth rate of battery management system market around world from 2021 to 2026, reprinted from [11], open access

EVs have been deployed. Besides, the growth in demand for consumer electronic products would further increase the demand for battery management, because battery management system is increasingly integrated into consumer electronic products for security purposes. In this context, effective and reliable battery management systems or solutions are urgently required to meet the requirements of these battery-based electronic products.

(2) **Battery management system basic functions**

In Li-ion batteries, the key to longevity, efficiency, reliability and safety lies in the efficient management of battery under various operating levels. For transportation electrification applications such as EV, the main basic functions of battery management include: battery data acquisition, battery modelling, battery states estimation, battery ageing prognostics, battery fault diagnosis, battery charging, etc. [12, 13]. In general, all battery management solutions first rely on the quality of collected battery data. The sampling speed, measurement accuracy, and data pre-filtering are initial key elements to determine battery management performance. In this context, battery management system typically requires various types of sensors to measure the data of battery current, voltage and temperature. Furthermore, several battery internal state information such as the state of charge (SoC), state of power (SoP), and state of health (SoH) are difficult to be measured directly; therefore, various filtering and estimation algorithms such as Kalman filter and its variants, particle filter (PF), and neural network (NN) are employed to obtain information of these states [14, 15]. Besides, battery ageing information such as further capacity degradation trajectory and remaining useful life need to be predicted for reducing EV users' mileage anxiety [16]. Another high priority of battery management is fault diagnosis to ensure battery safety, which means that any critical failures must be detected or battery system must be shut down if a fault occurs [17]. To achieve fast, safe, and efficient charging management, battery charging strategies with the ability to handle various conflicting objectives and satisfy battery operating constraints also need to be carefully designed [18]. The current battery management system mainly

monitors and controls batteries with fixed structures, which cannot provide full play to the optimal performance of battery systems. Han et al. designed a reconfigurable battery management system to allow the dynamic battery reconfiguration [19]. Dai et al. proposed a three layers-based battery management framework, including the foundation layer, algorithm layer and application layer [20]. These advanced battery management solutions are able to significantly improve battery safety, performance, and efficiency under various transportation electrification applications.

(3)   **Battery management system challenges**

Nowadays, due to transportation electrification being the broadest application scenario for Li-ion battery, battery management solutions are mainly designed for EV applications where the battery capacity ages from 100 to 80%. To ensure effective battery performance under complex, volatile, and extreme operating cases, various battery operation management strategies have been designed to protect electrical vehicle battery against faulty operations and to optimize its charging or discharging dynamics. However, in comparison with battery operation management area with fruitful solutions, fewer works have been done so far on applying data science-based strategies to benefit battery manufacturing and reutilization.

For battery manufacturing, as battery initial performance would be directly determined by each intermediate stage within the manufacturing line, an effective battery management solution that can analyse the effects of manufacturing parameters on battery properties and optimize the manufacturing line is crucial. Besides, the battery could make up to 30% weight and cost of an EV [21], while contributing to more than 40% CO_2 emissions during the production of EV [22]. In light of this, efficient management of battery manufacturing towards high-quality battery and economic targets such as high manufacturing yield, low manufacturing cost, and less pollution is crucial and plays a pivotal role in the acceptance of battery. Currently, as battery manufacturing generally contains a number of chemical, mechanical, and electrical operations, and also generates numerous strongly coupled manufacturing parameters in the order of tens or hundreds, engineers often rely on the experiment experience, expert advice, trial and error solutions to analyse and manage their battery manufacturing line. These solutions would result in huge laborious and time consumptions, slow battery product development, inaccurate quality control, and difficulty in generating sustainable business cases for the technological introduction. Therefore, it is imperative to introduce advanced and smart solutions to manage battery manufacturing, and explore the correlation, interaction, interdependency of all relevant parameters, to improve battery manufacturing performance.

For battery utilization, on the one hand, a Li-ion battery is usually determined to be unsuitable for EV applications when its real capacity becomes less than 80% of its nominal value [23]. As a result, a large number of automotive batteries will be retired in the coming years [24]. For example, 250,000 metric tons of automotive batteries are predicted to hit their end-of-life (EoL) by 2025 [25]. The second-life battery has the potential to generate more than 200 GWh by 2030, with a global value of more than $30 billion, according to another report [26]. Again, even under the most optimistic estimates, 3.4 million kg of automotive batteries cells might end up in the waste

stream by 2040 [27]. These numerous retired batteries containing volatile chemical elements would be released into the atmosphere if reutilization is not performed which will undergo both environmental and economic harm. On the other hand, in response to global climate change, many renewable and sustainable energy sources such as solar and wind have been adopted. However, due to the intermittent and time-varying existence of renewable energy sources, power fluctuates would be generated. This would significantly affect the grid performance, voltage stability, and reliability, making them become difficult to be processed into the grid. Based upon suitable battery reutilization solutions, this can be effectively mitigated if the generated energy from a renewable source is first deposited in a battery, and then converted by an appropriate power electronic converter topology to achieve the necessary grid voltage and frequency. In this context, giving such retired batteries a suitable reutilization solution, which is the management of batteries after they have reached 80% capacity would not only support the economy but also help to minimize total battery demand, resulting in a substantial reduction in the use of extracted chemical materials and significantly benefit many battery second-life applications such as grid energy storage [28].

Based upon the above discussions, battery full-lifespan from manufacturing, operation, and reutilization as a whole need to be carefully managed. With the rapid development of artificial intelligence and machine learning technology, data science-based tools stand out as the promising solutions for battery full-lifespan management, hopefully enabling us to overcome the major challenges dealing with different types of data from battery manufacturing, operation and reutilization. On the basis of this, a brand-new hologram to make full use of battery during full-lifespan could be formulated, further boosting the advancement of low-carbon technologies.

1.3 Data Science Technologies

To move data science-based tools applied to battery full-lifespan management efficiently, the systematic understanding and exploration of data science technology are required. In this context, data science-based tools must be properly explained and discussed in a way suitable for a broad audience.

1.3.1 What is Data Science

Data science is a practice of mining raw datasets with both structured and unstructured forms to identify specific patterns and extract meaningful insights from these data. It belongs to an interdisciplinary field, which mainly involves statistics, automation and engineering, computer science, machine learning, and new data-based technology to obtain insights from real data.
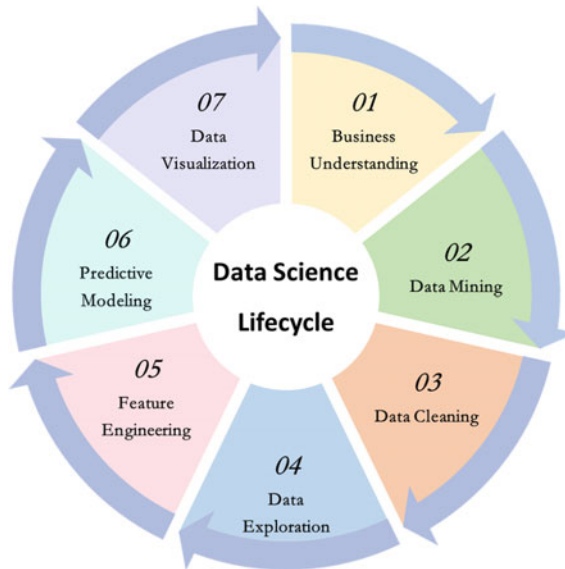
Fig. 1.12 Typical data science lifecycle

Figure 1.12 illustrates the typical lifecycle of data science, which includes seven main parts: business understanding, data mining, data cleaning, data exploration, feature engineering, predictive modelling, and data visualization. Business understanding mainly refers to the definition of relevant questions and objectives from the applications that require to be explored. For data mining, the necessary data needs to be gathered and scraped. Data cleaning involves the solutions to fix the inconsistencies within data and handle their missing values. For data exploration, data would be analysed visually to form the hypotheses of defined data science problem. For feature engineering, the importance and correlations of feature variables from data would be quantified and analysed. For predictive modelling, data science tools such as the machine learning models would be trained, validated, and adopted to make new predictions. For data visualization, conclusions will be reported to key stakeholders through various plots and interactive visualization tools.

In order to define data science task and clearly manage data science-based project, four main stages need to be carefully considered as:

Data architecture: The first stage in the data science pipeline workflow is to define data architecture. This requires data scientists to think through in advance how data users could make full use of data. Then data scientists also need to think about how to organize data to support different analyses and visualizations.

Data acquisition: The next stage is data acquisition which focuses on how to collect data from different sources such as experiments or real applications. Besides,

various representing, transforming, and grouping solutions are all needed to help data scientists understand how the data could be represented before analysis.

Data analysis: Data analysis is the core stage during data science workflow. In this stage, data scientists would use various technical, mathematical, and statistical tools such as AI and machine learning to conduct exploratory and confirmatory analysis works such as classification, regression, predictive analyses, and qualitative analyses.

Insight conclusion: After data analysis, data scientists would communicate the obtained insights through data visualization and reporting in this stage. These could benefit the stakeholders to obtain useful conclusions, readjust their strategies, and generate new plans for evaluation again.

In addition, numerous data science technologies need to be involved in the data science pipeline workflow. It should be known that all these data science technologies are designed by the programming language. Nowadays, the widely utilized programming language mainly includes:

(1) **Python**: Based upon some specifical and easy-to-implement libraries, Python becomes a popular open-source language, which has been widely used by academia and industries particular in AI community. After being created in 1989, Python became a feasible programming language to offer numerous tools for manipulating datasets and analyse data science results easily and conveniently [134]. Besides, there exist some Python libraries that specifically focus on machine learning algorithm development, including and not limited to Scikit-Learn, Keras, and TensorFlow. Due to numerous programming language forums and websites having published many topics on the implementation of Python, popularity becomes another merit of Python [131, 132]. In general, Python-based data science solutions need to be executed under a cross-platform named integrated development environment (IDE) including Pycharm, Spyder, and Jupyter Notebook, where the friendly interface and the possibility of interacting Python with other programming languages become key elements that need to be considered.

(2) **MATLAB**: As an efficient programming language for technical computing, MATLAB is able to integrate computation, visualization, and programming in an easy-to-implement condition where data science issues and approaches could be expressed in the familiar mathematical notations. The basic data element of MATLAB is just an array without the need for dimensioning, further benefitting the computational effort of numerous data science computing issues particular for those with the formulation of matrix or vector. After being developed over a period of years, MATLAB has become a standard instructional tool for several introductory and advanced courses in automation engineering and data science particular in an academic environment. Besides, MATLAB features a family of toolboxes for users to explore and apply different technologies for their specific applications. For data science, many toolboxes such as the neural network toolbox, deep learning toolbox, optimization toolbox,

and statistics and machine learning toolbox have been widely used to solve particular classes of issues.

(3) **R language**: After being developed around the last decade of the twentieth century, R language also becomes particularly popular in the field of statistics science. In comparison with Python and MATLAB, R language is less applied to develop data science solutions. However, it is able to provide fully dedicated statistical libraries including the MASS, stats, fdata, and glmnet. Besides, users could conveniently search the details regarding how to apply each library and package based on the Comprehensive R Archive Network (CRAN).

(4) **C++ and Fortran**: C++ and Fortran belong to the modern pioneers of programming languages, which have been widely adopted as high-performance language. For the data science applications, several C++ libraries including the SHARK and MLPACK can be used to design machine learning. In comparison with Python, MATLAB, and R language, the implementation of data science code through C++ and Fortran would become more difficult as the necessary memory management is usually required.

1.3.2 Type of Data Science Technologies

Data science technologies mainly contain the types of supervised, unsupervised, and semi-supervised approaches. For the supervised ones, certain variables need to be defined as input and output terms before a related dataset is employed. On the contrary, the definition of input and output terms is missing for the unsupervised data science methods, whose target is to automatically discover patterns in the dataset. Semi-supervised methods are somewhere in between supervised and unsupervised ones, which would apply datasets containing both labelled and unlabelled data.

Supervised data science approaches could be further divided into two main categories including regression methods and classification methods. The data science regression model analyses and outputs data in terms of continuous values, while the data science classification model will analyse and output classes in terms of discrete values. These classes utilized for supervised data science approaches can come from the operator or from the unsupervised data science methods. For battery management applications, apart from the type adopted, classical data science approaches heavily depend on the data and are quite unrelated to physics, which means that they can be aimed at, for example, determining the underlying mapping among various variables, rather than providing any physical explanation of such mapping. However, through coupling battery physical elements, physical-driven data science methods also exist.

Figure 1.13 shows several most utilized data science methods in battery research and development. It should be known that all these methods have been adopted in the applications of battery full-lifespan management.

Neural network (NN): As illustrated in Fig. 1.13a, NN is proposed to mimic human brain activities through using the processing unit of artificial neurons arranged in the input layer, output layer as well as hidden layers. After a pre-processing stage,
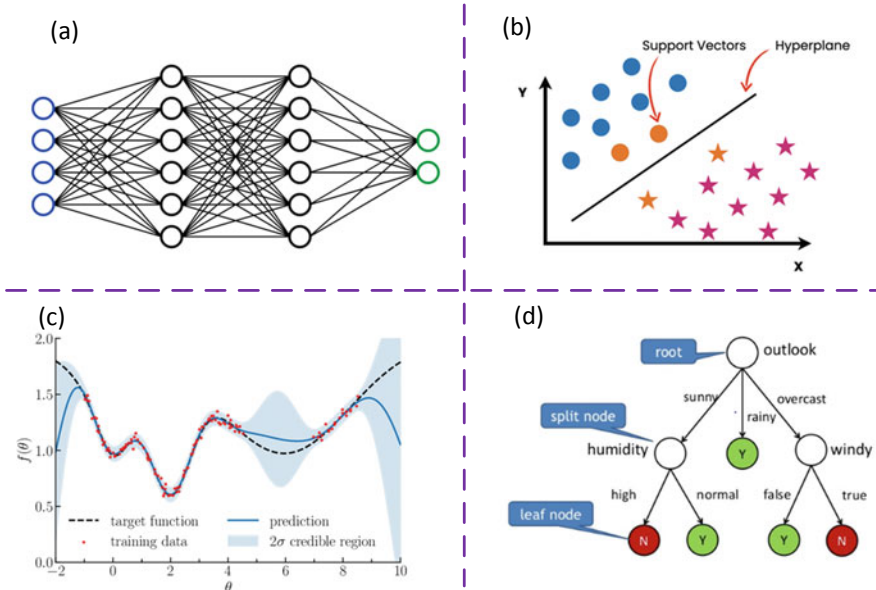
Fig. 1.13 Several typically utilized data science methods in battery research and development: **a** neural network, **b** support vector machine, **c** Gaussian process regression, **d** tree-based solution

data will be inputted into the input layer with a predefined input matrix. Then the neurons in hidden layers contain the mathematical function to generate output across neurons and could be expressed through using a weighted linear combination being wrapped in the activation functions [29]. In theory, The larger the neuron weight, the sensitivity to this specific input would become higher. Finally, the output layer would output the predicted values of NN. For the NN training process, the parameters here are mainly optimized by considering the amount of hidden layers, the number of neurons within each layer, the interconnected neuron weights, and the activation function types.

To date, two different types of NNs are widely utilized in battery management applications, including the feedforward-neural network (FNN) and recurrent-neural network (RNN). In the former, the data would travel in one direction only. After equipping feedback connection with FNN, RNN is derived. By involving the recurrent links, RNN is capable of keeping and updating the previous information for a period of time, making it a promising tool to capture the sequential correlations in battery management applications. For example, as battery ageing process usually contains hundreds of cycles, while the ageing information among these cycles is highly correlated. It is thus meaningful to extract and store these correlations for accurate battery lifetime prognostics. Besides, as it is able to capture the long-term dependency of data, RNN thus becomes an effective tool for capturing and updating sequential information during battery management.

One obvious benefit of NN is its capability of studying from experience and can adapt to varied situations. However, NN requires a large amount of battery application data to train and verify, and its accuracy would be heavily influenced by the training way and data quality. Furthermore, the computational effort is still a bottleneck for its large-scale application in battery management and NN structure also plays a pivotal role in determining its performance. In this context, NN optimization still remains an open technical problem. In general, the NN structure is determined by time-consuming trial and error. In light of this, some optimization approaches such as the two-stage stepwise identification method [30] could be adopted to optimize the NN structure for battery management applications.

Support vector machine (SVM): According to the kernel functions, SVM belongs to a supervised data science tool. It could perform both classification and regression tasks by searching the hyperplane separating classes with a maximal margin, as shown in Fig. 1.13b. SVM could adopt kernels to handle nonlinear problems by transforming the nonlinear issue with a low-dimensional space into a linear problem with a higher-dimensional space [31]. In theory, the prediction of SVM is based on several functions defined over the input space, and learning is a process to infer this function's parameters. SVM could make predictions with the function as:

$$y(x) = \sum_{n=1}^{N} \omega_n K(x, x_n) + \varepsilon \tag{1.1}$$

where ω_n represents the weights to connect feature space into output, $K(\cdot)$ stands for kernel function, and ε represents the independent noise.

SVM is particularly appealing for its ability to handle training datasets with small size. Here the number of support vectors would increase when the size of the training dataset becomes larger. To enhance the stability and robustness of SVM under large-scale training dataset size, decremental and incremental solutions [32] could be adopted by integrating relevant data samples for SVM training and ignoring irrelevant parts. However, the computational effort would be also increased in this process.

Gaussian process regression (GPR): Deriving from the Bayesian framework, GPR-based data science models have been widely adopted in battery prognostic applications due to their superiority in terms of being flexible, nonparametric, and probabilistic [33]. GPR is also a kernel-based data science approach, which is capable of realizing predictions combined with prior knowledge as well as providing variance around its mean prediction point to express the associated uncertainties, as shown in Fig. 1.13c. Here the Gaussian process can be regarded as the collection of a limited amount of random variables that present the joint multi-variate Gaussian distribution. In theory, the performance of GPR is significantly sensitive to the kernel functions, so the kernel functions need to be carefully designed for achieving high prediction accuracy. Battery application is usually complicated and would be affected by many impact elements. The single kernel function will easily lead to unreliable predictions for nonlinear mappings with multi-dimensional input terms.

In this context, an isotropic kernel with advanced structures such as the automatic relevance determination could be utilized. Furthermore, hyperparameters optimization of kernel functions within GPR is crucial as improper hyperparameters are easy to lead overfitting issue. To ameliorate this, minimizing the negative log marginal likelihood is generally adopted [34].

Tree-based solutions: Tree-based solutions are the decision-support data science by adopting the flowchart-like model to achieve classification or regression, as illustrated in Fig. 1.13d. Many tree-based solutions such as decision tree (DT), random forest (RF), and boosting-based approaches have been successfully utilized in the applications of battery management. The basic idea of DT is to divide a complicated prediction issue into many smaller ones based on a tree structure. In this way, each node within a DT could represent a small subissue, while DT as a whole could constitute a solution to the overall issue [35]. For DT training, data would be first injected in a root (i.e. the first node of DT). After that, the input term which could best discriminate between the output would be searched. That is, which value (V_i) of which input terms could split the initial dataset in such a way to separate as many outputs as possible would be searched to minimize related errors. This would result in the node being divided into two paths: one for values of the selected input items lower than V_i while another for values larger than V_i. The iteration of this process would result in a series of paths linking possible inputs to a certain output. One obvious benefit of DT is that an easy-to-understand representation of the links between input and output items could be generated. However, due to the too simple structure, DT is difficult to achieve high performance particularly for highly nonlinear applications. To handle this, many ensemble learning solutions such as RF are designed through combining DTs to improve overall prediction performance. The logic of RF can be summarized as that if the single DT cannot provide results with enough accuracy, the result through averaging all outputs from numerous DTs with a bagging solution would result in more accurate predictions [36]. This could bring significant improvement of prediction performance, further making RF become competent to solve highly nonlinear issues. Besides, boosting-based approaches also adopted some DTs to decrease both bias and variance of derived data science models, while the related prediction accuracy could be improved. The main difference here is the adopted sampling approach, where bagging and boosting solutions are differentiated by the procedure utilized for the training process.

1.3.3 Performance Indicators

After establishing data science-based solutions for different battery management applications, a key task is to adopt suitable performance indicators for evaluating the performance of these data science-based solutions. These performance indicators can be divided into two main categories to evaluate the regression and classification results, respectively.

(1) **Data science regression model**

To quantify and evaluate the accuracy of devised data science-based regression models in various battery management cases, the following three typical performance indicators are usually utilized.

Mean absolute error (MAE): Supposing N is the total number of regression samples, Y_i represents the actual reference value while $\hat{Y}_i$ stand for the output predicted from data science regression models, then the MAE could be obtained to evaluate the repression accuracy as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| Y_i - \hat{Y}_i \right| \tag{1.2}$$

Root mean square error (RMSE): According to the same character definition, RMSE is another typical performance indicator to present the deviations between the predicted output and actual reference value as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2} \tag{1.3}$$

R^2 value: Supposing $\overline{Y}_i$ is the mean value of all response outputs, R^2 value is also a typical performance indicator to reflect how closely the outputs from regression model could match well with the actual reference values as:

$$R^2 = 1 - \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2 / \sum_{i=1}^{N} \left(Y_i - \overline{Y}_i \right)^2 \tag{1.4}$$

For the regression applications, when the outputs predicted from models get close to the real experimental ones, MAE and RMSE present to be close to 0, while R^2 would get close to 1, indicating that a data science regression model is capable of explaining all the variability of target outputs.

(2) **Data science classification model**

For the classification cases, to quantify and evaluate the performance of the designed data science classification model, several performance indicators including the confusion matrix, macro-precision, macro-recall, and macro-F1-score are generally adopted.

Precision rate (P_{rate}): Supposing positive corresponds to the class of interest while negative corresponds to other classes, four basic measures including the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) could be formulated for each class. Then for the class of interest C_i ($i = 1, \ldots, N_c$, N_c is the

number of classes), its P_{rate} could be obtained to quantify the correct classification result of this class as:

$$P_{\text{rate}} = \text{TP}/(\text{TP} + \text{FP}) \tag{1.5}$$

Recall rate (R_{rate}): R_{rate} is able to quantify the rate of all fraud cases of this class as:

$$R_{\text{rate}} = \text{TP}/(\text{TP} + \text{FN}) \tag{1.6}$$

F-measure (F-measure): F-measure can reflect the harmonic mean of precision as well as recall of this class as:

$$F\text{-measure} = \frac{2 \times P_{\text{rate}} \times R_{\text{rate}}}{P_{\text{rate}} + R_{\text{rate}}} \tag{1.7}$$

Overall correct classification rate (OCC_{rate}): OCC_{rate} that reflects the proportion of correctly classified observations out of all the observations could be obtained by:

$$\text{OCC}_{\text{rate}} = \frac{\text{TP}_{\text{all}} + \text{TN}_{\text{all}}}{N} \tag{1.8}$$

where $\text{TP}_{\text{all}} + \text{TN}_{\text{all}}$ stands for all the correctly classified outputs from data science classification model, N represents the total amount of observations.

Confusion matrix (CM): According to the aforementioned metrics, a CM with $M+1$ rows and $M + 1$ columns can be formulated to reflect the performance of multiple class-based classification model. Here each row within CM is able to reflect the predicted output classes while each column stands for the actual target classes. The elements on the primary diagonal of CM reflect the correctly classified results, while other elements stand for the incorrectly classified conditions. The $M + 1$th column and $M + 1$ th row stand for the $P_{\text{rate}}(C_i)$ and $R_{\text{rate}}(C_i)$ of each class, respectively. The last element in the right-bottom corner is the OCC_{rate}.

macro$P$, **macro$R$**, and **macro$F$1**: Supposing each class has a $P_{\text{rate}}(C_i)$, $R_{\text{rate}}(C_i)$, and F-measure (C_i), then various overall performance indicators including the macro-precision (macro$P$), macro-recall (macro$R$), and macro-F1-score (macro$F$1) could be obtained to evaluate the overall classification performance of data science classification model as:

$$\begin{cases} \text{macro}\,P = \sum_{i=1}^{N_c} P_{\text{rate}}(C_i)/N_c \\ \text{macro}\,R = \sum_{i=1}^{N_c} R_{\text{rate}}(C_i)/N_c \\ \text{macro}\,F1 = \sum_{i=1}^{N_c} F\text{-measure}(C_i)/N_c \end{cases} \tag{1.9}$$

For the classification applications, when the classes outputted from a model match observations as much as possible, $\text{macro}\,P$, $\text{macro}\,R$, and $\text{macro}\,F1$ would get close to 1, indicating that the data science classification model is able to perform high accurate classification.

Based upon the aforementioned classical performance indicators, the performance of data science-based battery management solutions can be quantified and evaluated.

1.4　Summary

This chapter first introduces the background and motivation of Li-ion battery. It outlines the role of Li-ion battery in the energy storage market of several leading countries. Three applications to comprise the bulk of the current Li-ion battery market including the electric vehicle, electronic device, and stationary battery-based energy storage are also introduced. Then, it describes the fundamental of Li-ion battery and the demands of battery management. Apart from battery operation management with fruitful solutions, the management of both battery manufacturing and reutilization is still in its infancy. In this context, with the rapid development of artificial intelligence and machine learning, data science-based solutions become a promising way to handle various key challenges of battery full-lifespan management. After that, this chapter reviews the basic information on data science lifecycle and widely utilized programming language and outlines the popular data science technologies used in battery full-lifespan management and corresponding performance indicators for result evaluation. It emphasizes the necessity and benefits of using data science technologies to manage batteries, while also guiding the design and development of data science-based tools for effective battery full-lifespan management.

References

1. China Energy Storage Alliance: Energy Storage Industry White Paper 2021. http://en.cnesa.org/
2. Guiding Opinions on Accelerating the Development of New Energy Storage (Draft for Comment). http://www.nea.gov.cn/139896047_16230559585841n.pdf
3. National Energy Administration: http://www.nea.gov.cn/2021-07/30/c_1310097494.htm
4. Frazier AW, Cole W, Denholm P, Greer D, Gagnon P (2020) Assessing the potential of battery storage as a peaking capacity resource in the United States. Appl Energy 275:115385

5. Golombek R, Lind A, Ringkjøb H-K, Seljom P (2022) The role of transmission and energy storage in European decarbonization towards 2050. Energy 239:122159
6. HM Government Industrial Strategy—The road to Zero, 2018
7. SMMT UK Automotive. www.smmt.co.uk/industry-topics/uk-automotive/
8. Martin N, Rice J (2021) Power outages, climate events and renewable energy: Reviewing energy storage policy and regulatory options for Australia. Renew Sust Energ Rev 137:110617
9. Large-scale battery storage. Australian Renewable Energy Agency. Knowledge Sharing Report. ARENA, 2019. https://arena.gov.au/assets/2019/11/large-scale-battery-storage-knowledge-sharing-report.pdf
10. National Blueprint for Lithium Batteries 2021–2030. https://www.energy.gov
11. Mordor Intelligence,: Battery management system market- growth, trends, Covid-19 impact, and forecasts (2021–2026). https://www.mordorintelligence.com/industry-reports/battery-management-system-market
12. Lu L, Han X, Li J, Hua J, Ouyang M (2013) A review on the key issues for lithium-ion battery management in electric vehicles. J Power Sources 226:272–288
13. Liu K, Li K, Peng Q, Zhang C (2019) A brief review on key technologies in the battery management system of electric vehicles. Front Mech Eng 14(1):47–64
14. Hu X, Feng F, Liu K, Zhang L, Xie J, Liu B (2019) State estimation for advanced battery management: key challenges and future trends. Renew Sust Energ Rev 114, 109334
15. Wang Y, Tian J, Sun Z, Wang L, Xu R, Li M, Chen Z (2020) A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. Renew Sust Energ. Rev 131:110015
16. Hu X, Xu L, Lin X, Pecht M (2020) Battery lifetime prognostics. Joule 4(2):310–346
17. Hu X, Zhang K, Liu K, Lin X, Dey S, Onori S (2020) Advanced fault diagnosis for lithium-ion battery systems: a review of fault mechanisms, fault features, and diagnosis procedures. IEEE Ind Electron Mag 14(3):65–91
18. Tomaszewska A, Chu Z, Feng X, O'kane S, Liu X, Chen J, Ji C, Endler E, Li R, Liu L (2019) Lithium-ion battery fast charging: a review. ETransportation 1:100011
19. Han W, Wik T, Kersten A, Dong G, Zou C (2020) Next-generation battery management systems: dynamic reconfiguration. IEEE Ind Electron Mag 14(4):20–31
20. Dai H, Jiang B, Hu X, Lin X, Wei X, Pecht M (2021) Advanced battery management strategies for a sustainable energy future: multilayer design concepts and research trends. Renew Sust Energ Rev 138:110480
21. Nykvist B, Nilsson M (2015) Rapidly falling costs of battery packs for electric vehicles. Nat Clim Change 5(4):329–332
22. Duffner F, Kronemeyer N, Tübke J, Leker J, Winter M, Schmuch R (2021) Post-lithium-ion battery cell production and its compatibility with lithium-ion cell production infrastructure. Nat Energy 6(2):123–134
23. Winslow KM, Laux SJ, Townsend TG (2018) A review on the growing concern and potential management strategies of waste lithium-ion batteries. Resour Conserv Recycl 129:263–277
24. Lybbert M, Ghaemi Z, Balaji A, Warren R (2021) Integrating life cycle assessment and electrochemical modeling to study the effects of cell design and operating conditions on the environmental impacts of lithium-ion batteries. Renew Sust Energ Rev 144:111004
25. Stringer D, Ma J (2018) Where 3 million electric vehicle batteries will go when they retire. Bloom. Businessweek 27
26. Lovell J. Storage: retirement home for old EV batteries? https://www.energycouncil.com.au/analysis/storage-retirement-home-for-old-ev-batteries/
27. Casals LC, García BA, Canal C (2019) Second life batteries lifespan: rest of useful life and environmental analysis. J Environ Manage 232:354–363
28. Deng Y, Zhang Y, Luo F, Mu Y (2020) Operational planning of centralized charging stations utilizing second-life battery energy storage systems. IEEE Trans Sustain Energy 12(1):387–399
29. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: a survey. Heliyon 4(11):e00938

30. Li K, Peng J-X, Bai E-W (2006) A two-stage algorithm for identification of nonlinear dynamic systems. Automatica 42(7):1189–1197
31. Suthaharan S (2016) Machine learning models and algorithms for big data classification. Springer, Berlin, pp 207–235
32. Chen Y, Xiong J, Xu W, Zuo J (2019) A novel online incremental and decremental learning algorithm based on variable support vector machine. Clust Comput 22(3):7435–7445
33. Schulz M, Speekenbrink A, Krause A (2018) A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. J Math Psychol 85:1–16
34. Lucu M, Martinez-Laserna E, Gandiaga I, Camblong H (2018) A critical review on self-adaptive Li-ion battery ageing models. J Power Sources 401:85–101
35. Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227
36. Altman N, Krzywinski M (2017) Ensemble methods: bagging and random forests. Nat Methods 14(10):933–935