EFFAT UNIVERSITY
COLLEGE OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING (ECE)

Online Prediction of the Li-Ion Battery Remaining Useful Life for Smart Grid by Using Event-Driven Approach

A thesis submitted in partial fulfillment for the degree of Master's in Science in Energy Engineering (MSEE) per the requirements of Effat University

By
Amal Essam ElDin AbdelGawad

Supervised by
Dr. Saeed Mian Qaisar
Associate Professor & Researcher, Electrical & Computer Engineering (ECE) Department
College of Engineering, Effat University, Jeddah, KSA

and
Dr. Muhammed Akbar
Associate Professor, Natural Science, Math, and Technology Unit (NSMTU)
College of Engineering, Effat University, Jeddah, KSA

Jeddah, KSA
May 2021 – Shawwal 1442

جامعة عفَّت

كلِّية الهندسة

قسم الهندسة الكهربائية وهندسة الحاسبات

التنبُّؤ عن بُعد بالعمر الإنتاجي المتبقِّي لبطَّارية أيون الليثيوم للشبكة الذكية باستخدام النهج القائم على الحدَث

أطروحة مقدَّمة لِنَيل درجة الماجستير في العلوم في هندسة الطاقة حسَب متطلَّبات جامعة عفَّت

إعداد

أمل عصام الدين عبد الجوَّاد

إشراف

د. سعيد ميان قيصر

أستاذ مشارك وباحث، قسم الهندسة الكهربائية وهندسة الحاسبات

كلِّية الهندسة، جامعة عفَّت، جدَّة، المملكة العربية السعودية

ود. محمَّد أكبر

أستاذ مشارك، وحدة العلوم الطبيعية والرياضيات والتكنولوجيا

كلِّية الهندسة، جامعة عفَّت، جدَّة، المملكة العربية السعودية

جدَّة، المملكة العربية السعودية

شوَّال ١٤٤٢هـ – مايو ٢٠٢١م


بسم الله الرحمن الرحيم

APPROVAL PAGE

Effat University

Jeddah, Kingdom of Saudi Arabia

Deanship of Graduate Studies and Research

This thesis, written by Amal Essam ElDin AbdelGawad under the direction of her thesis supervisor and approved by her thesis committee, has been presented to and accepted by the Dean of Graduate Studies and Research on May 24th, 2021, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE in Energy Engineering – Renewable Energy Concentration.

Thesis Committee

Thesis Supervisor
Name: Dr. Saeed Mian Qaisar.

Signature: .....

Co-supervisor
Name: Dr. Muhammed Akbar.

Signature: .....

Department Chair
Name: Dr. Enfel Barkat.

Signature: .....

External Member
Name: Dr. Essam A. Al-Ammar.
Title: Associate Professor.

Signature: .....

Dean of the College
Name: Dr. Akila Sarirete.

Signature: .....

Internal Member
Name: Dr. Tayeb Brahimi.
Title: Assistant Professor.

Signature: .....

Dean of Graduate Studies & Research
Name: Dr. Mady A. Mohamed.

Signature: .....

صفحة الإجازة

جامعة عفت

جدة، المملكة العربية السعودية

عمادة الدراسات العليا والبحث العلمي

كَتبت هذه الرسالة الطالبة أمل عصام الدين عبد الجوَّاد، تحت إشراف المشرف المكلَّف بالإشراف على رسالتها، وأجازتها لجنة التحكيم، وقُدِّمت إلى عميدة الدراسات العليا والبحث العلمي بجامعة عفَّت، بصفتها جزءًا من متطلَّبات الحصول على درجة الماجستير في العلوم، برنامج هندسة الطاقة – تخصُّص هندسة الطاقة المتجدِّدة، وقد جرت الموافقة على الرسالة وأُجيزَت بتاريخ ١٢ شوَّال ١٤٤٢هـ، الموافق ٢٤ مايو ٢٠٢١م.

أعضاء لجنة التحكيم

المشرف على الرسالة

الاسم: د. سعيد ميان قيصر.

التوقيع: .....

المشرف المشارك

الاسم: د. محمَّد أكبر.

التوقيع: .....

رئيسة القسم

الاسم: د. أنفال بركات.

التوقيع: .....

العضو الخارجي

الاسم: د. عصام عبد العزيز العمَّار.

التوقيع: .....

عميدة الكلية

الاسم: د. عقيلة سريرات.

التوقيع: .....

العضو الداخلي

الاسم: د. طيِّب براهيمي.

التوقيع: .....

عميد الدراسات العليا والبحث العلمي

الاسم: د. ماضي أحمد محمَّد.

التوقيع: .....

DECLARATION

I hereby declare that this thesis titled **"Online Prediction of the Li-Ion Battery Remaining Useful Life for Smart Grid by Using Event-Driven Approach"** is based on my original work except for quotations and citations been duly acknowledged. I also declare that the proposed dissertation has not been previously or concurrently submitted for the award of any degree at Effat University, any other university or institution.

Student Name: Amal Essam ElDin AbdelGawad.

Signature: 

Date: May 18th, 2021.

ACKNOWLEDGMENT

First of all, I thank Allah for providing me with the ability and strength to finish this work. Secondly, I would like to thank all the amazing people who supported me throughout my master's thesis. It is a delight to show gratefulness to those who made the completion of this thesis possible.

I expand my blessings to my family, without whom finishing the master's program has been a far-off reality—an incredible appreciation to my parents and siblings for their trust, supports, and endless love throughout my life. Indeed, I genuinely appreciate Dr. Saeed Mian Qaisar, my supervisor, whose support and guidance were significant. I owe my deepest gratitude to him for his supervision, special attention, and time that he gave me, which genuinely helped smooth the work. I would also like to express my most profound appreciation to Dr. Mohammad Akbar, my co-supervisor, for his continuous support and encouragement.

Additionally, I would like to express my respect to Dr. Aziza Ibrahim, the former ECE Department Chair. Her support is broadly acknowledged during my master's program. Lastly, I would like to thank the evaluation committee for both the two semesters of my work, Dr. Enfal Barakat, Dr. Nema Salem, Dr. Tayeb Brahimi, and Dr. Essam Al-Ammar, who gave me their time to evaluate my thesis and provide me with valuable comments to improve my performance better.

DEDICATION

To my grandfathers' souls.

ABSTRACT

The latest innovative advancements have included renewable energy sources based on smart grids and electric vehicles (EVs). Technologies and equipment related to remote and integrated power systems offer improved developments for smart and microgrids. The battery is a crucial element of modern power systems; it is indispensable in many vital applications such as EVs, drones, avionics, satellites, mobile phones, and energy storage for renewable smart grids.

Due to their superior characteristics, including but not limited to having a compact size, high power supply capability, and a higher number of charging/discharging cycles, the Lithium-ion (Li-ion) batteries are one of the most dominant energy storage technologies. On the other hand, due to one of their most significant disadvantages being expensive, and to optimize their performance and ensure they last longer in the smart grid, their use is monitored using battery management systems (BMSs). The extensive processing resources that modern BMSs need can result in higher overhead power consumption. Regarding this, several embedded and integrated systems-based solutions have been proposed. This thesis focuses on upgrading the present Li-ion BMSs, reconstructing their associative data acquisition by employing the event-driven sensing mechanism. It aims at efficiently predicting the Li-ion battery's remaining useful life (RUL) online using machine learning (ML).

Firstly, a suitable current high-power Li-ion battery real dataset has been distinguished and utilized during offline processing. Then, this dataset is reconstructed to make its form compatible with MATLAB. Using MATLAB, the proposed system is modeled. The event-driven peak sensing model efficiently extracted the features from the studied battery consumption parameters based on the phenomenon of shape context feature extraction. The feature extractor is sufficiently achieved via embedding the event-driven peak sensing phenomenon in the system. These features are used for training the proposed event-driven remaining useful life (RUL) predictor using robust ML classifiers, namely k-Nearest Neighbor (kNN), Artificial Neural Network (ANN), Linear Regression (LR), Random Tree (RT), and Random Forest (RF). During online processing, the event-driven sensing-based acquired parameters are used to predict the battery capacity. Besides, by applying an event-driven peak sensing approach, the shape context features are extracted. Beyond, those fused features are used for online RUL prediction of the studied Li-ion battery cell by employing Weka software.

The devised solution excelled its counterparts in terms of power consumption with a compression gain of 437.5-fold on average. The results showed superior performance for the LR among its studied counterparts, followed by the ANN, kNN, RF, then RT. The achieved MAE, RMSE, RAE, and RRSE by the LR based on the training and testing split of 70:30% are 0.0019, 0.0023, 1.0411%, and 1.1606%, respectively, and by the 5-fold cross-validation are 0.00289, 0.0051, 1.6405%, and 2.6656%, respectively.

The proposed solution is parametrizable and can be utilized in different expected applications such as smart and microgrids, hybrid electric vehicles (HEVs), drones, distributed sensors, and satellites.

Keywords: Event-Driven Processing; Li-Ion Battery; Battery Management System; Remaining Useful Life; Battery Capacity; Data Compression; Feature Extraction.

الملخَّص

تضمَّنت أحدثُ التطوُّرات المبتكَرة توزيعَ مصادر الطاقة المتجدِّدة القائمة على الشبكات الذكية، والسيَّارات الكهربائية. وتُقدِّم تقنياتُ ومعدَّاتُ أنظمة الطاقة اللاسلكية والمتكاملة تطوُّراتٍ محسَّنةً للشبكات الذكية والشبكات الصغيرة. في هذا السياق، نجد أنَّ البطَّارية عنصر حاسم في أنظمة الطاقة الحديثة، إذْ لا غنى عنها في كثير من التطبيقات الحيوية، مثل المركبات الكهربائية والطائرات المسيَّرة وإلكترونيات الطيران، والأقمار الصناعية والهواتف المحمولة، وتخزين الطاقة في الشبكات الذكية.

تُعدُّ بطَّاريات أيون الليثيوم إحدى أقوى تقنيات تخزين الطاقة هيمنةً، نظرًا لخصائصها الممتازة، بما في ذلك على سبيل المثال لا الحصر: الحجم الصغير، والقدرة العالية على الإمداد بالطاقة، والعدد الأكبر من دورات الشحن والتفريغ. لكن نظرًا لواحد من أكبر عيوبها المتمثِّل في كونها باهظة الثمن، وسعيًا إلى تحسين أدائها وضمان استمرار عملها ضمن الشبكة الذكية لمدَّة أطول، فإنَّ استخدامها يُراقَب باستخدام أنظمة إدارة البطَّارية. يمكن أن تؤدِّي موارد المعالجةِ المكثَّفةِ التي تحتاج إليها أنظمة إدارة البطَّارية إلى زيادة استهلاك الطاقة العامَّة. وفي هذا الصدد، اقتُرِحَت عدَّة حلول قائمة على الأنظمة المدمَجة والمتكاملة.

تركِّز هذه الأطروحة على ترقية أنظمة إدارة بطَّارية أيون الليثيوم بواسطة إعادة تصميم كيفية الحصول على بياناتها الترابطية، بتوظيف آلية المعالجة القائمة على الحدَث. وتهدف إلى تقدير العمر الإنتاجي المتبقِّي لبطَّارية أيون الليثيوم عن بُعد بكفاية، باستخدام خوارزميات تعلُّم الآلة.

أولًا، في أثناء المعالجة في وضع عدم الاتصال، جرى اختيار مجموعة بيانات حقيقية مناسبة لبطَّارية أيون الليثيوم الحالية عالية الطاقة، ثم عُولِجَت مجموعة البيانات هذه لجعل شكلها متوافقًا مع برنامج MATLAB، الذي باستخدامه صُمِّم النظام المقترح. بعد ذلك، استخرَج نموذجُ استشعار الذروة الميزاتِ من محدِّدات استهلاك البطَّارية المدروسة، اعتمادًا على ظاهرة استخراج ميزات سياق الشكل. حُقِّقَت خطوة استخراج الميزات تحقيقًا فعَّالًا بتضمين ظاهرة استشعار الذروة القائمة على الحدَث في النظام. واستُخدِمَت هذه الميزاتُ لتدريب النموذج المقترَح للتنبُّؤ بالعمر الإنتاجي المتبقِّي استنادًا إلى الحدَث، باستخدام مُصَنِّفَات تعلُّمِ آلةٍ قوية، وهي: خوارزمية الجار الأقرب (kNN)، والشبكة العصبية الاصطناعية (ANN)، والانحدار الخطِّي (LR)، والشجرة العشوائية (RT)، والغابة العشوائية (RF). في أثناء المعالجة عن بُعد، استُخدِمَت المحدِّدات المكتسَبة المستنِدة إلى الاستشعار القائم على الحدَث، للتنبُّؤ بقدرة البطَّارية. إلى جانب ذلك، وبتطبيق نهج استشعار الذروة القائم على الحدَث، استُخرِجَت ميزاتُ سياق الشكل. ثم استُخدِمَت هذه الميزاتُ المدمَجة للتنبُّؤ عن بُعد بالعمر الإنتاجي المتبقِّي لخلية بطَّارية أيون الليثيوم المدروسة، باستخدام برنامج Weka.

لقد تفوَّق الحلُّ المبتكَر على نُظَرائه من حيث استهلاك الطاقة، مع مكسب انضغاط بلغ في المتوسِّط ٤٣٧٫٥ ضعفًا. لقد أظهرت النتائج أداءً فائقًا للانحدار الخطِّي (LR) بين نظرائه المَدرُوسين، تَلَتْه الشبكة العصبية الاصطناعية (ANN)، فالجار الأقرب (kNN)، فالغابة العشوائية (RF)، ثم الشجرة العشوائية (RT). وقد حقَّق الانحدار الخطِّي باستخدام نسبة ٧٠٪ لمجموعة التدريب و٣٠٪ لمجموعة الاختبار متوسِّطَ خطأٍ مُطلَقٍ (MAE) بقيمة ٠٫٠٠١٩، وجذرَ متوسِّطِ خطأٍ تربيعيٍّ (RMSE) بقيمة ٠٫٠٠٢٣، وخطأً مطلقًا نسبيًّا (RAE) بقيمة ١٫٠٤١١٪، وجذرَ خطأٍ تربيعيٍّ نسبيٍّ (RRSE) بقيمة ١٫١٦٠٦٪. وكانت قيمة كلٍّ منهم على التوالي باستخدام التصديق المتقاطع لخمس تقسيمات: ٠٫٠٠٢٨٩ و٠٫٠٠٥١ و١٫٦٤٠٥٪ و٢٫٦٦٥٦٪.

إنَّ الحل المقترَح قابل للتحديد ويمكن استخدامه في مختلِف التطبيقات المتوقَّعة، مثل الشبكات الصغيرة والذكية، والمركبات الكهربائية الهجينة، والطائرات المسيَّرة، وأجهزة الاستشعار الموزَّعة، والأقمار الصناعية.

الكلمات المفتاحية: المعالجة القائمة على الحدث، بطارية أيون الليثيوم، نظام إدارة البطارية، العمر الإنتاجي المتبقي، قدرة البطارية، ضغط البيانات، استخراج الميزات.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADC	Analog-to-Digital Converter
ANN	Artificial Neural Network
AR	Autoregressive
BMS	Battery Management System
CO_2	Carbon Dioxide
CV	Cross-Validation
DM	Delta Modulation
DoD	Depth of Discharge
DV	Differential Voltage
ECM	Equivalent Circuit Model
ED	Event-Driven
EDADC	Event-Driven Analog-to-Digital Converter
EDS	Event-Driven Sensing
EIM	Electrochemical Impedance Model
EIS	Electrochemical Impedance Spectroscopy
EKF	Extended Kalman Filter
EM	Electrochemical Model
EoD	End of Discharge
EoL	End of Life
EV	Electric Vehicle
G2V	Grid-to-Vehicle
GCC	Gulf Cooperation Council
GHG	Greenhouse Gas
GPM	Gaussian Process Mixture
GPR	Gaussian Process Regression
HEV	Hybrid Electric Vehicle
HPPC	Hybrid Pulse Power Characterization
IC	Incremental Capacity
kNN	k-Nearest Neighbor

Li-ion	Lithium-ion
LOCR	Linear Optimizing Combination Resampling
LR	Linear Regression
LSB	Least Significant Bit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
NiMH	Nickel-Metal Hybrid
NMC	Lithium Nickel Manganese Cobalt Oxide
OCV	Open-Circuit Voltage
OIES	Oxford Institute for Energy Studies
PAM	Pulse Amplitude Modulation
PCM	Pulse Code Modulation
PCoE	Prognostics Center of Excellence
PF	Particle Filter
PS	Percentage Split
PSO	Particle Swarm Optimization
RAE	Relative Absolute Error
RF	Random Forest
RMSE	Root Mean Squared Error
RRSE	Root Relative Squared Error
RT	Random Tree
RUL	Remaining Useful Life
RVM	Relevance Vector Machine
SAMA	Saudi Central Bank (Saudi Arabian Monetary Authority, formerly)
SCPF	Spherical Cubature Particle Filter
SDG	Sustainable Development Goal
SEI	Solid Electrolyte Interface
SMC	Sequential Monte Carlo

SNR	Signal-to-Noise Ratio
SoC	State of Charge
SoC-OCV	State of Charge Open-Circuit Voltage
SOE	Safe Operating Envelope
SoH	State of Health
SoH_E	The state of Health-related to Energy capacity
SoH_P	The state of Health-related to Power capacity
SoL	State of Life
SVM	Support Vector Machine
SVR	Support Vector Regression
UKF	Unscented Kalman Filter
UN	United Nations
UPF	Unscented Particle Filter
V2G	Vehicle-to-Grid
WPME	Wiener Process with Measurement Error

LIST OF SYMBOLS

C	The least significant bit (LSB) or the comparator step
c_i^{act}	Actual capacity for the i^{th} instant
$\overline{c^{act}}$	Mean of actual capacity
c_i^{pred}	Predicted capacity for the i^{th} instant
$\overline{c^{pred}}$	Mean of predicted capacity
dt_n	Difference between the present and previous sampling instant
k	Number of the cycle or iteration
M	Total number of cycles that have been considered
n	Total number of the instants or response values
N	Number of bits per sample in the convertor
$N_{classic}$	Number of samples acquired in the classical counterpart
N_{ED}	Number of samples acquired in the suggested model
P_{Noise}	Average noise power
P_{Signal}	Average signal power
$Q_{current}$	Actual cell capacity at the current time
$Q_e(t)$	Quantization error
Q_{fresh}	Initial cell capacity when it was fresh
r	Correlation coefficient
R^2	Decision coefficient
R_{Comp}	Compression ratio
$s(t)$	Sampling function
t	Time
t_{ED_k}	The k^{th} instant when the cell parameter reaches the preset threshold in the suggested model
TH_b	The b^{th} threshold
t_n	Present sampling instant
t_{n-1}	Previous sampling instant
T_s	Sampling interval
v	Testing set size

V_{C_k}	The k^{th} cell terminal voltage while considering the charging cycles
V_{D_k}	The k^{th} cell terminal voltage while considering the discharging cycles
$x(t)$	Analog signal
$\tilde{x}_n(t)$	Analog signal in the case of EDADC
x_n	Present sampling amplitude
x_{n-1}	Previous sampling amplitude
$\tilde{x}_{n-1}(t)$	Previous analog signal in the case of EDADC
$x_Q[n]$	The discrete quantized signal at a given n
$x_Q(t)$	Quantized version of an analog signal
$x_{ref}(t)$	Digital reference signal in the case of EDADC
$x_s(t)$	The ideal sample version of an analog signal
δ	Impulse

CHAPTER 1
INTRODUCTION

1.1. Overview

The high economic growth testified in the Gulf Cooperation Council (GCC) region has been co-occurred by a high annual population growth rate of about 3% on average between 2010 and 2020; according to United Nations (UN), the GCC population increased from 45,101,978 in 2010 to 58,664,095 in 2020 [1]. This rapidly growing population requires significant investment in infrastructure and services and raises important questions and pressure on energy demand and supply, so governments, not only in the GCC but also worldwide, are thinking about the best way to meet future energy needs. According to the Saudi Central Bank (SAMA), the total annual electric energy consumption of Saudi Arabia in 2020 was 289.328 TWh [2], the highest consumption among all GCC countries. The Oxford Institute for Energy Studies (OIES) has stated that, based on energy consumption trends in Saudi Arabia after witnessing economic and industrial developments in the past decades, projections indicate that domestic energy consumption will increase by 4% to 5% annually until 2030 [3]. This growth is not only unsustainable in terms of resources, but the biggest problem is what causes it to increase the proportion of greenhouse gas (GHG) emissions that cause substantial climate change and global warming and damage the natural environment, which is mainly caused by the use of fossil fuels such as coal, oil, and gas. That clearly shows why the world is now moving towards renewable energy. By 2030, the GHG emissions rate that the GCC countries will avoid is expected to reach between 1 and 50 million tons of CO_2 [4]. Furthermore, achieving the goals of Saudi Arabia's Vision 2030 will contribute to mitigating the climate, fulfilling the environmental sustainability plans, and enhancing economic growth [5].

In pursuit of the 7^{th} and 11^{th} Sustainable Development Goals (SDGs) of the Paris Agreement signed in September 2015, which aim to "ensure access to affordable, reliable, sustainable, and modern energy" and "make cities and human settlements inclusive, safe, resilient and sustainable," respectively [6], renewable energy sources are widely used. There has been a fast development in the usage of renewable energy during the past decade. Among various renewable resources, Saudi Arabia has great natural potential for wind and solar power [7]. These renewable energy resources are

intermittent; thus, they should be incorporated with storage equipment to be distributed over a smart grid. By using energy storage technologies, power flows can be organized, and their fluctuations can be reduced. Moreover, the incorporation of renewable energy production resources can be reinforced, as energy storage units represent backup units that store excess energy production when the generated energy is greater than the energy demand and discharge it at peak times. That is why the importance of renewable energy storage in smart grids has increased, as it provides a practical solution for higher network controllability and power stability-enhancing [8]. It is the perfect partner for renewable energy.

Another perspective is that urban emissions exacerbate health concerns due to traditional road transportations. Therefore, the production of low carbon and oil-independent transport solutions is seen as essential. In Saudi Arabia, for example, road transportation accounts for a sizeable GHG share that harms the environment [9]. That is why the world moves towards electric vehicles (EVs) as an optimal alternative to internal combustion engine vehicles and as a booster for smart grid upgrade as it is based on the principle of bidirectional communication, which is enhanced by the concept of Vehicle-to-Grid (V2G) and Grid-to-Vehicle (G2V) [10].

EVs use various energy storage devices such as lead-acid, nickel-metal hydride, and lithium-ion (Li-ion) batteries. Li-ion batteries are widespread due to their high energy and power density, no memory effect [11], long service life, and very high efficiency. Because of the Li-ion battery's advantageous characteristics, many investments have already been made to improve its stability and reliability. Despite the high initial cost, the Li-ion battery market's growth is steadily growing, which is expected to continue. If the Li-ion battery can be improved to work more safely and reliably in EVs, gasoline usage may be overcome, reducing the GHG effects [12]. Big agencies invest in methods that can accurately estimate every cell's state in the battery pack and, thus, predict its remaining useful life (RUL). Here we can realize the importance of this thesis topic: Li-ion battery RUL predicting.

1.2. Problem Statement

Due to one of the most significant disadvantages of Li-ion batteries being expensive, and to optimize their performance and ensure they last longer in the smart grid, their use is monitored using battery management systems (BMSs). However, the

extensive processing resources that modern BMSs need can result in higher overhead power consumption. Therefore, this thesis focuses on upgrading existing Li-ion BMSs by enhancing data acquisition mechanisms to achieve real-time data compression. It aims at predicting the remaining useful life (RUL) of Li-ion batteries to provide more safety and reliability for them.

1.3. Importance and Motivation

According to Saudi Arabia's Vision for 2030, Saudi Arabia seeks to move from fossil fuels as a source of electricity and utilize renewable energy such as solar and wind [13]. As a part of Vision 2030 goals, it was announced in the last quarter of 2017 that the city of NEOM will be established in the northwest of Saudi Arabia. The basic concept that NEOM represents is being a city that integrates smart cities technologies and applies the most efficient renewable energy sources to power itself with only green energy [14].

In that context, energy storage technologies are vital because of the intermittency and instability of renewable energy [8]; they will be the key source of power. Furthermore, the smart grid helps balance electricity demand and supply by digitally changing the intended transmission line's power load, depending on its demand [15], producing a strong continuous supply of energy throughout the hour.

Many advantages make Li-ion batteries at the forefront of the best and most mature energy storage solutions, including but not limited to having light weights, high power density, and long cycle life [16]. Despite that, these batteries are costly; thus, the battery management systems (BMSs) are utilized regularly to efficiently use them and ensure that they are proper for various applications in the smart grid. That will contribute to decreasing the risk of failure of Li-ion batteries that leads to vast blackouts. Moreover, monitoring the remaining useful life (RUL) in Li-ion battery packs, which is the time between a Li-ion battery's current and failure condition, is necessary to enhance the battery performance and guarantee minimum degradation.

However, the extensive processing resources that modern BMSs need can result in higher overhead power consumption [17]. Regarding this, several embedded and integrated systems-based solutions have been proposed. Nevertheless, there is a need to upgrade existing Li-ion BMSs and enhance data acquisition and prediction mechanisms for the Li-ion batteries' RUL.

1.4. Aim and Objectives

This thesis aims to reduce overhead usage; the purpose is to enhance present state-of-the-art battery management systems (BMSs) efficiency. The objectives are to reconstruct the current Li-ion BMSs associative data acquisition by employing the event-driven sensing mechanism; to efficiently predict the Li-ion battery's remaining useful life (RUL) online, in pursuit of a safer and more reliable Li-ion battery in the smart grid.

1.5. Outline

The following parts of the thesis include four chapters, the list of references, and appendices. In Chapter 2, the researcher studies the background and main literature review and research on Li-ion batteries, battery management system (BMS), state of charge (SoC), state of health (SoH), and remaining useful life (RUL). Chapter 3 discusses the research methodology in detail, including the system's block diagram, the dataset description, battery parameters sensing, feature extraction, machine learning-based RUL prediction, models evaluation techniques, and performance evaluation metrics. Chapter 4 illustrates the thesis results in both tabular and graphical representations and discusses them. Finally, Chapter 5 summarizes what was mentioned in the thesis and mentions the scientific contribution, some suggested future works, and recommendations. The five chapters are followed by two appendices, one for publications and another for some tabular results.

CHAPTER 2
BACKGROUND AND LITERATURE REVIEW

This chapter of the thesis presents its background on Li-ion batteries, battery management system (BMS), state of charge (SoC), state of health (SoH), and different recent reviews and important studies on remaining useful life (RUL) estimations for Li-ion batteries, presented by several excellent works, publications, and research related to the investigated topic.

2.1. Lithium-Ion (Li-Ion) Batteries

There are different types of batteries with various applications. Batteries can be either primary or secondary batteries. Primary batteries are disposable, while secondaries are rechargeable. The most mature secondary batteries are Lithium-ion (Li-ion) batteries [18]. They were first designed and developed by Asahi Kasei during the 1980s and commercialized by Sony and A&T Battery Corporation in 1991 and 1992, respectively, which has rapidly grown ever since. A few years after commercialization, Li-ion batteries took over half of the mobile electronic devices market. The technology is developing rapidly for EVs and stationary electrical energy storage applications [19], [20].

In the past, only nickel-cadmium batteries were compatible with mobile devices until the 1990s came, and nickel-metal hybrid (NiMH) and Li-ion batteries appeared, which was a technological breakthrough. Nowadays, Li-ion is the fastest growing and most promising battery technology. Li-ion technology is widely used in consumer low-power products in the field of electronics such as cell phones and laptops and high-power applications such as EVs, trams, and bicycles. Li-ion batteries have become a common energy source, and their performance continues to improve so they can be used in applications that require more power. Li-ion batteries have superior attributes compared to NiMH batteries, lead-acid, nickel-cadmium, or nickel-metal hydride cells. They have lighter weights and higher specific energy density (23-70 Wh/kg), very high efficiency (near 90%), relatively long cycle life (3000 cycles at 80% depth of discharge), high nominal voltage, no memory effect, and less maintenance requirement. Due to these enhanced attributes, we currently can find the Li-ion batteries within our day-to-day used electronics [11], [16], [20]–[22].

Moreover, high energy and power density make the Li-ion batteries compatible with hybrid and electric vehicles (HEVs and EVs). Besides, its high efficiency makes them usable in various applications, such as improving the quality of renewable energies like wind and solar. Thus, the previous advantages make the Li-ion batteries active contributors to renewable energies on a wide scale [12].

However, despite all these advantages, Li-ion batteries have some disadvantages. The first is their high price compared to lead-acid batteries, for example. The most famous type of Li-ion battery is the lithium nickel manganese cobalt oxide (NMC), a combination of $LiCoO_2$ (LCO), $LiMn_2O_4$ (LMO), and nickel; thus, this type of Li-ion battery consists of NMC and graphite. This mixture of NMC is the battery's cathode (positive electrode), and the graphite is the anode (negative electrode). When including a simple battery management system (BMS) and casing to calculate the cost price of a Li-ion battery pack, its estimate would be around 300-432 $/kWh. The main reason for that high cost is that the battery's two primary and more expensive components, the cathode and anode, require large amounts of noble materials. Moreover, if including the profit margins of the manufacturer and distributor, the selling price is up to 670 $/kWh for low production quantities. The cost breakdown of the Li-ion battery with a special focus on the anode composition is represented in Figure 2.1. With the reach of 2025-2030, it is expected that the prices will be lower, reaching 100 $/kWh for NMC-based Li-ion batteries [23].

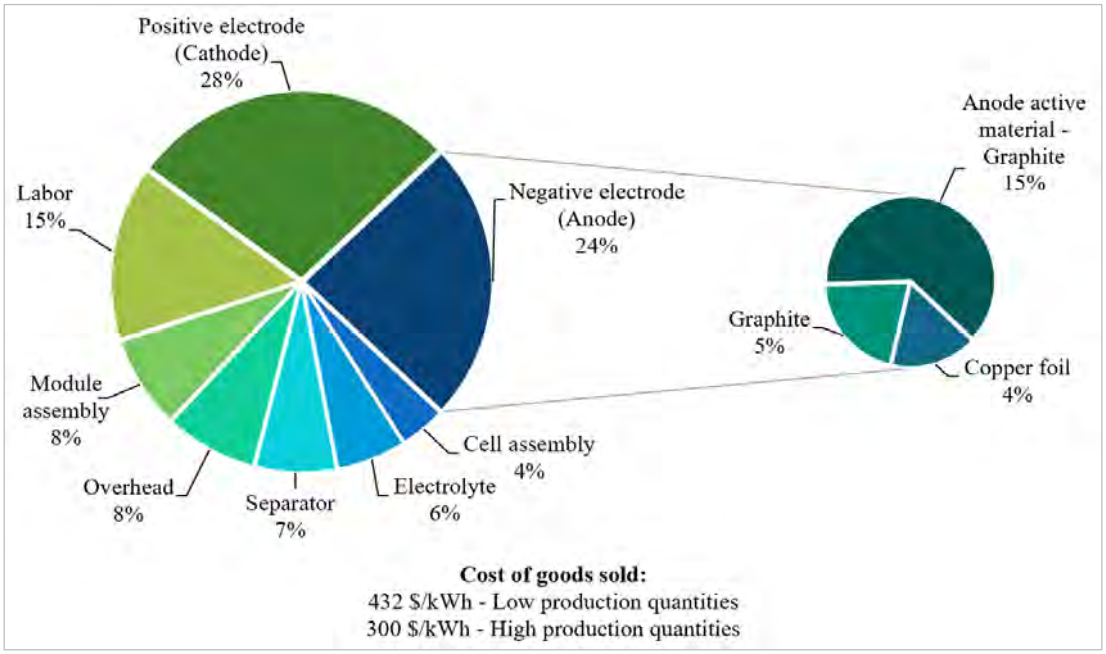


Figure 2.1: Li-ion battery cost breakdown with a special focus on the anode composition [23]

Like all batteries, Li-ion batteries go through degradation. The degradation is a kind of aging in which the battery's materials deteriorate over its lifetime, so its performance degrades through operation over time. Every time the battery goes through charging/discharging cycles, the degradation process accelerates, making the battery's efficiency, power, and capacity further affected [21]. This degradation could lead to unwanted events ranging from inconvenience, financial loss, or fatal tragedies. Therefore, it is necessary to assess and determine a battery's expected lifetime [20], [24]. For instance, it was reported that Sony and Samsung had lost significantly financially due to their Li-ion batteries failures in their phones and laptops. Sony lost around $430 million due to recalling 9.6 million laptops globally [20], while Samsung has endured a higher loss of $17 billion after the release of Galaxy Note 7 due to its multiple explosions recorded [16].

Nowadays, most Li-ion batteries utilize graphite-based carbonaceous anodes whose operating voltage is around 0.05V. Compared to traditional liquid organic electrolytes whose stabled voltage ranges from 1V to 4.5V, the graphite-based anodes are considered theoretically unstable. During the charging/discharging operations, a passivation protective layer is constructed on the anode electrode's surface due to the Li-ion consumption and electrolyte depletion. This layer is commonly known as the solid electrolyte interface (SEI) layer. This layer is typically created during the first few charging cycles, particularly in the first charging cycle. That results in the deterioration in the capacity of the Li-on battery. As the battery continues charging/discharging, the anode size would vary by around 10% due to intercalation and deintercalation of lithium ions. This variation in the volume may lead to cracking of the SEI layer, which results in an interaction between the electrolyte and lined graphite. Consequently, the inventory of electrolyte and lithium ions is rapidly consumed. Furthermore, the SEI layer is continuously constructed, and the capacity of the battery is further reduced. The internal resistance of the battery is also increased. The increase in SEI thickness can be considered as one of the critical factors responsible for Li-ion batteries aging [25].

In general, there are several concerns related to the utilization and storage of Li-ion batteries. These concerns include the cell electrolyte leaking, internal gas venting, fires, and explosion of battery cover exposing interior components. Overheating of the Li-ion batteries can occur for a variety of reasons. It could be due to a short circuit or inadequate storage condition concerning the following facts. As the

temperature rises, the self-discharge rate of the cell increases, resulting in reduced battery performance; thus, the cell lifespan or remaining useful life (RUL) is reduced. On the other hand, the electrolyte can evaporate at high temperatures, increasing the cell pressure, which may cause a fire or explosion. In light of that, one of the most serious consequences of uncontrolled overheating is the release of combustible heat and gas. In order to avoid battery explosions, some cells design has been modified to include a specific vent to release the gases. However, such vents might be obstructed or incorrectly open in some cases. That results in the battery expanding, container ruptures, and eventually, battery exploding. On the other hand, if such specific vents do not exist in the cells, such as pouch cells, the gases may be discharged at weak places in the exterior pouch [26]. Consequently, companies and researchers have turned their attention to evaluate the reliability of batteries to avoid unfortunate incidents [16], [20], [22], [24], [27].

To avoid abuse and extend the lifespan by monitoring and controlling Li-ion batteries' performance, the battery management system (BMS) has already been developed and widely implemented [21]. Indicators such as state of charge (SoC); an estimator of the charge capacity remaining post or during usage to indicate the timing of battery recharge, state of health (SoH); a predictor of the discharge capacity that would provide an estimation of the battery age, and remaining useful life (RUL); an indicator to the remaining life of a battery until its end of life (EoL). These indicators or estimators represent the monitoring part of the BMS [16], [20], [22], [24], [27], where this thesis works.

2.2. Battery Management System (BMS)

This part of the thesis discusses the battery management systems (BMS) used in Li-ion battery packs. An efficient BMS is imperative for the Li-ion battery to work safely and reliably, prevent any physical damage, and deal with thermal degradation and cell imbalance. Moreover, various battery conditions, such as SoC and SoH, can be assessed with an efficient battery management system, detecting temperature, measuring voltage and current, and adjusting alarms to avoid overcharging/over-discharging. Besides, BMS has significant importance for data monitoring and updating, fault detection, battery voltage equalization, which are crucial factors to achieve good SoC and SoH accuracy [12].

A BMS has two primary goals for a battery pack: keeping the battery pack operating safely and working reliably. All the advanced features found at a BMS come down to one of these two features at its core. These features are intended to achieve the best battery pack performance. As shown in Figure 2.2, the BMS measures and monitors the voltage, current, and temperature of each cell in the battery pack, and controls the cooling, balancing, and power limits of the battery pack [12], as well as operates the disconnect breakers for start-up and power-off.

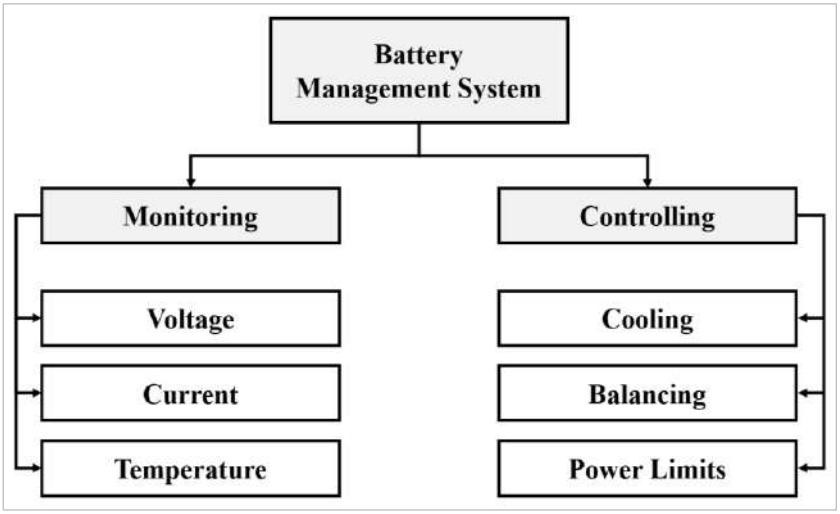


Figure 2.2: BMS system main functions

Let us look at the black-box model of a BMS in Figure 2.3. We can notice that at a high level and its core, the BMS is taking some inputs and running some algorithms and estimators to produce several outputs to the application system controller. For example, the BMS takes several voltages, such as every cell voltage and pack voltage. It also takes, as an input, the current flowing into or out of the battery pack, i.e., the battery pack charging/discharging and at what magnitude. Moreover, the BMS uses several temperature sensors to get a sense of the cells' temperature distribution. Then, within the BMS, it runs many algorithms to generate an accurate estimation of the outputs. One of the primary outcomes is the state of charge (SoC) estimation. That is commonly referred to as a fuel gauge and is the standard charge, showing up on the phone or EV as a remaining charge. The BMS also generates a state of health (SoH) estimation. At a high level, SoH represents how much capacity the battery pack can store relative to when it first got it or the beginning of life. In its second year, the battery pack is always not as good as its first year. Here, an accurate prediction of a battery's remaining useful life (RUL) is essential in intelligent battery health management

systems [28]. The BMS is also developing a state of the safe operating envelope (SOE), which allows us to know how much current can be either charged or discharged at any given time. Finally, the BMS is outputting some faults or status signals that the application controller needs to be aware of, and so, there can be certain conditions where these are triggered [29].

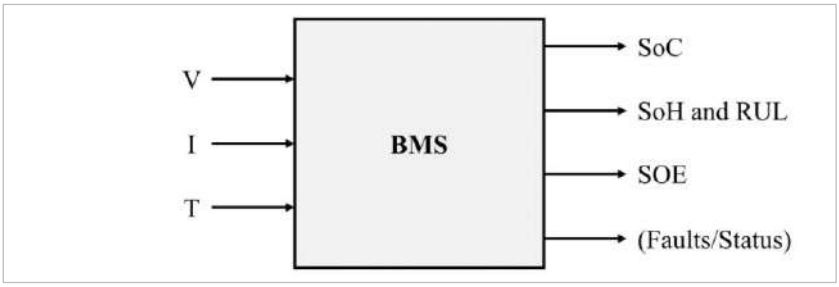


Figure 2.3: BMS system block diagram

In a schematic, a BMS can be used in single or multi-cell Li-ion battery applications. Figure 2.4 shows a 3S-1P battery pack with three cells connected in series. The BMS is a circuit board that typically resides very close to the cells, monitors the voltage to measure each cell voltage, and monitors the overall battery pack voltage. Besides, the BMS measures the current flowing into or out the pack via a shunt sensor, for example. Finally, the BMS has a master disconnect that allows it to terminate the battery charge or discharge if it detects that the battery system is entering into an unsafe or undesirable state. Then, external terminals connect the battery pack to the load, i.e., an EV, a solar power inverter, or a DC-DC converter. In a nutshell, the BMS monitors everything going on internally and communicates via an interface to an external controller that uses this information to update its behavior better [29], [30].

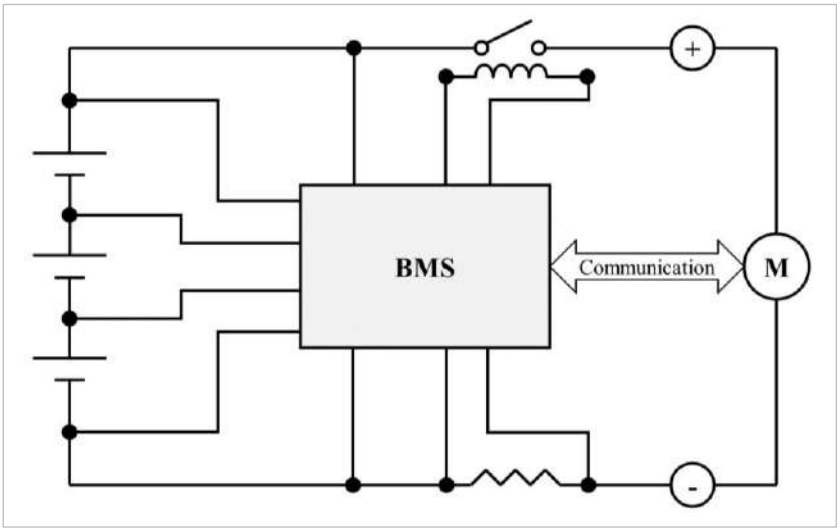


Figure 2.4: BMS schematic and communication diagram

2.3. State of Charge (SoC) Estimation

In this part of the thesis, the state of charge (SoC) will be discussed, as used in Li-ion battery packs. SoC estimation for all energy storage devices has always been a big problem. A highly accurate SoC estimation not only gives us information on how much charge is left in the battery or the remaining usable energy but also estimates the reliability of the battery. Besides, an accurate and efficient SoC estimation gives an idea about charging/discharging strategies that significantly impact battery applications, where each cell can have different capacities due to aging, temperature, self-discharge, and manufacturing variance [12].

As mentioned above, a precise and efficient assessment of the SoC of a battery can provide the required assessment factor for vehicle energy management and the control system's optimal design. Therefore, more methods have been suggested for measuring the SoC of a battery in real-time. For a more detailed comparison of these methods, we divided them into four categories [31], and this classification is summarized in Figure 2.5.

The first SoC estimation method is the looking-up table-based method in which the SoC of batteries has a direct mapping relationship with their external (static) characteristic parameters, such as the battery open-circuit voltage (OCV) and AC impedance. This approach is more fitting for the laboratory environment. The second method is the Ampere-hour integral method when the maximum available battery capacity is known and its current can be measured precisely. The ampere-hour integral method can allow the accurate calculation of the variation of the SoC. This method is more successful when applied in combination with other techniques, such as model-based methods. The third type of SoC estimation method is model-based estimation methods, which can be summarized into three categories: electrochemical model (EM), equivalent circuit model (ECM), and electrochemical impedance model (EIM). The last SoC estimation method is the data-driven estimation method that is a control method that uses the system's input-output data to construct a controller. Under this category, there are two types of SoC estimation methods. The first is the data training method that shows highly accurate results. Its algorithms are still susceptible to their parameters when the training set cannot fully cover the current operating conditions. In contrast, the other method is the data-model fusion method, representing a method that combines

the online data-driven method and the model-based method, thus ensuring the system's convergence and stability [31].

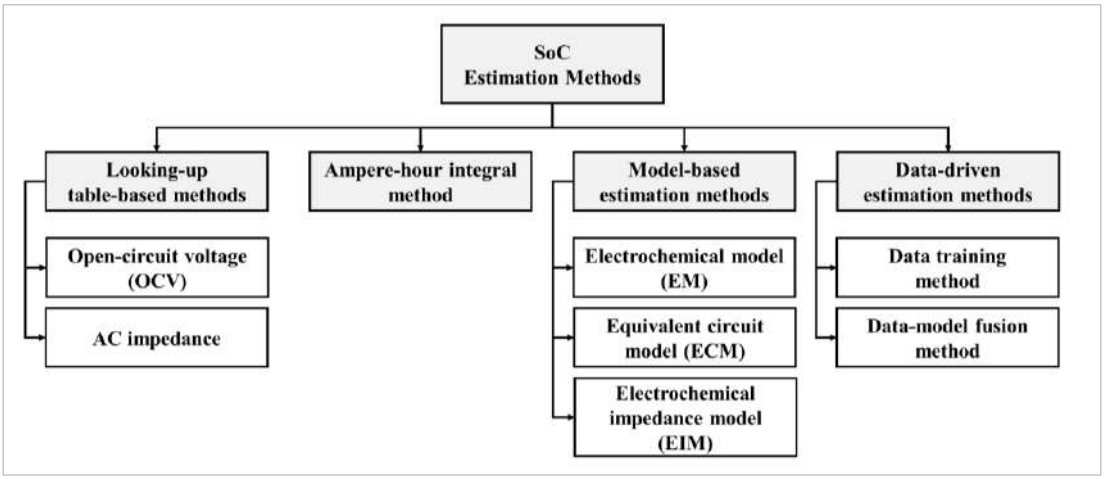


Figure 2.5: Types of SoC estimation methods [31]

2.4. State of Health (SoH) Estimation

In this part of the thesis, the state of health (SoH) estimation will be discussed, as used in Li-ion battery packs. The battery SoH represents a battery's ability to store and supply energy concerning its initial conditions by considering the application's energy and power requirements, which can be described as a state of health-related to energy capacity (SoH_E) or/and power capacity (SoH_P). The battery capacity can measure the SoH_E, and impedance is used to describe the battery's SoH_P quantitatively. The protection of entire systems needs to provide a reliable prediction of either SoH_E or SoH_P with simple approaches. With this information, we can adjust the operating modes to extend the battery life and predict the appropriate time durations for battery replacement [32].

In [32], by calculating the ratio of actual cell capacity to the initial cell capacity, the battery capacity is chosen as the SoH indicator:

$$SoH = \frac{Q_{current}(Ah)}{Q_{fresh}(Ah)} \times 100\% \quad (1)$$

As shown in Equation (1), the SoH percentage can be calculated by calculating the ratio of the actual cell capacity at the current time ($Q_{current}$) to its initial capacity when it was fresh (Q_{fresh}). This value is equal to 100% for a fresh battery; then, it degrades over time.

The application requirements determine the battery end of life (EoL), representing that stage when the battery can no longer meet its performance criteria in terms of capacity or power. For example, batteries used in electric and hybrid electric vehicles (EVs and HEVs) expire and must be replaced when their SoH is less than 80%. However, this direct method of calculation given by Equation (1) is subject to the complete battery charging and discharging cycle. Therefore, this method is not practical in actual application because the batteries, in most cases, are not completely charged and discharged [32].

Recently, research on the SoH of batteries has garnered a lot of researchers' interest. In this field, many estimation techniques have been introduced. Each technique has different advantages and disadvantages regarding the accuracy of the estimation, the duration of the testing time, and the feasibility of implementation. As summarized in Figure 2.6, the SoH estimation techniques can be classified into three categories: adaptive models, experimental techniques, and incremental capacity / differential voltage analysis. The first category is the adaptive model-based methods, which can be further categorized into equivalent circuit-based models and electrochemical methods. These models have a strong physical relationship between the model parameters and the implied electrochemical processes in the battery cells. Unfortunately, model-based techniques can be computationally intensive and challenging to implement for actual implementation in BMS due to large matrix operations. The second category consists of experimental techniques divided into measurement-based methods and direct measurements from experiments, such as hybrid pulse power characterization (HPPC) and electrochemical impedance spectroscopy (EIS). The third category consists of differential methods based on incremental capacity (IC) analysis and differential voltage (DV) analysis. This differential analysis has the advantages of both experimental and adaptive-model methods, as it can be used for battery degradation identification as well as the SoH estimation with low computational effort. This method's main drawback is the requirement of completely static charging/discharging [32].

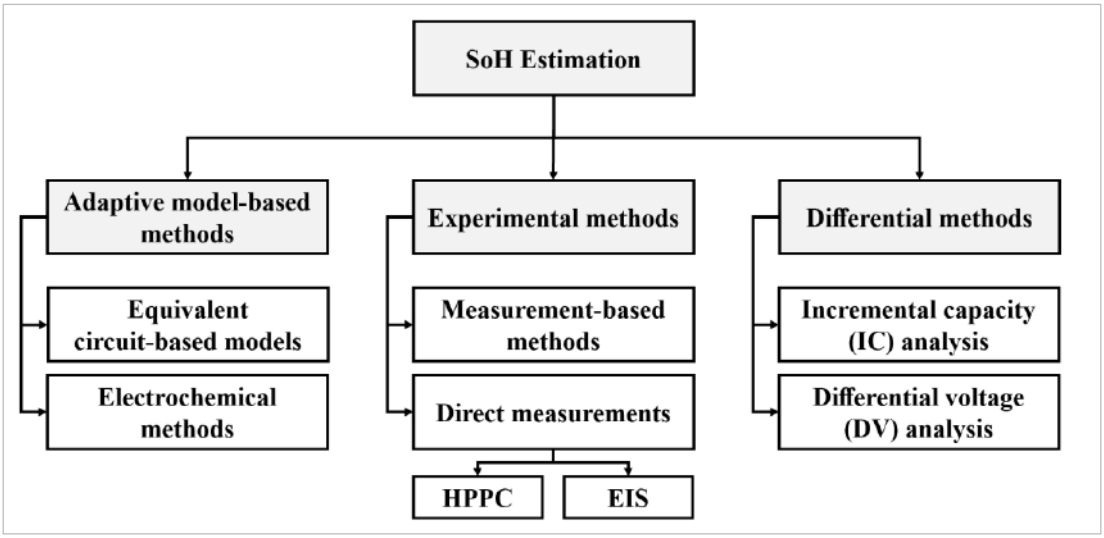


Figure 2.6: Types of SoH estimation methods [15]

The authors in [21] indicate that, to the present time, there is no unique optimal method for SoH estimation. The choice should be made based on what is to be estimated and available data. For example, in the case of possessing an enormous amount of data and a possible simple algorithm that can be used, the combination of degradation mechanism detection and significant data methodology would be the most suitable and sufficient. In another case, like obtaining data little by little as the tests are being developed, it would be more appropriate to use an adaptive model combined with the degradation.

2.5. Remaining Useful Life (RUL) Estimation

The remaining useful life (RUL) represents the time between a Li-ion battery's current and failure condition, as shown in Figure 2.7. This part of the thesis reviews recent studies on the RUL prediction for Li-ion battery packs.

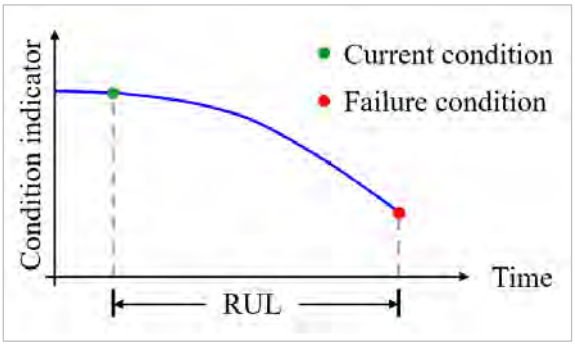


Figure 2.7: Decaying battery profile

Numerous researchers and organizations have increased their efforts to develop approaches and methods to accurately estimate a Li-ion battery's remaining useful life

(RUL). The particle filter (PF) method, or else known as the sequential Monte Carlo (SMC) method, is a method that combines Bayesian learning techniques with importance sampling. Qiang et al. [20] in 2013 proposed an improved version of the PF method. The author has developed the unscented particle filter (UPF) method, consisting of two steps. The first step uses the Unscented Kalman filter (UKF) algorithm to obtain the proposal distribution. The second step uses the PF algorithm to get the final results. Consequently, a degradation model was designed. The model and algorithms that ran prediction tests recorded an error percentage of less than 5% for actual RUL estimation for Li-ion batteries, which is better than the error percentage recorded using PF algorithms (around 7%) [20].

Zhang et al. [16] in 2018 managed to further improve the UPF method for the prediction of RUL of Li-ion battery by using linear optimizing combination resampling within the UPF (U-LOCR-PF); this was applied to overcome the particle diversity deficiency that exists in the previous methods. The author has pointed out that step coefficient K requires applying the fuzzy inference system as the determination of the coefficient K can significantly influence the performance of LOCR. RUL predictions showed improved accuracy when it is compared relative to the previous PF and UPF prediction methods. The authors achieved an average root mean squared error (RMSE) of 0.0335 and 0.0201 using 40 and 70 cycles.

Many researchers in the literature have used regression models to act as a predictive method for Li-ion batteries. As statistical methods, Long et al. [22] in 2013 proposed an improved autoregressive (AR) mode by utilizing the RMSE after stating that a uniform criterion that would provide AR order determination does not exist. The enhanced particle swarm optimization (PSO) algorithm was also applied to determine the best AR model order. Metabolism of the information within the data allowed the AR model order to change adaptively. It was concluded that this approach predicts the RUL of Li-ion batteries with a small margin of error and is also applicable for onboard applications. Hu et al. [33] in 2014 conducted similar research on rechargeable Li-ion batteries implanted in medical devices by testing their reliabilities with a developed non-linear kernel regression model. The developed model was based on the k-nearest neighbor (kNN) regression. The conducted research relied on charging voltage and current curves. It should be noted that the use of the kNN regression model allowed the inclusion of characteristics features of charging curves, which the author defined, and

the use of PSO decreased the cross-validation (CV) error in the capacity estimation. Finally, it was concluded that the proposed model has accurately predicted the life the of Li-ion battery for its complete life according to verification of 10 years of cycling data, achieving an RMSE of 1.08875 on average.

As data-driven methods based on the machine-learning, Wang et al. [34] in 2013 developed a Li-ion battery prognostic model consisting of a relevance vector machine (RVM) algorithm and a capacity degradation model for Li-ion batteries RUL prediction. The RVM selects the significant training vector; three conditional capacity degradation models are constructed to fit the relevance vectors' predictive values. The model is validated by three studies and has acceptable predictive outcomes. The authors achieved an RMSE of 0.0082 on average. Ismail et al. [35] in 2014 proposed an SVM-based iterative multi-step prediction model to attain accurate RUL prediction. Working temperatures and energy efficiency are used as inputs to characterize the training dataset. Experiments and reports evaluating the model showed high accuracy in detecting Li-ion batteries' RUL characteristics with few parameters.

Patil et al. [27] in 2015 presented another form of the regression model to predict the real-time RUL of Li-ion batteries. That was achieved by combining support vector regression (SVR) with a support vector machine (SVM), taking advantage of their classification and regression attributes. The model aimed to train an algorithm utilized in EVs to predict the battery near its end of life (EoL) and notify the driver before its complete depletion. The model was fed with cycling data of batteries under different circumstances extracted from the battery data set. Case studies were carried out for the model, and it was concluded that it could be used as a potential onboard RUL estimation tool for EVs. Zhang et al. [36] in 2018 introduced a recurrent neural network-based long short-term memory (LSTM) for Li-ion batteries' RUL prediction. Different Li-ion cells were used under different currents and temperatures to prepare the experimental data. The model obtains good results in predicting RUL independently of the offline training data.

Miao et al. [37] in 2013 developed a unified particle filter-based degradation model to predict the RUL for Li-ion batteries. The proposed model achieves better accuracy than the PF method in RUL prediction with less than 5% error. Tang et al. [38] in 2014 developed a novel prognostic method with the Wiener process with measurement error (WPME) for RUL prediction. Fang et al. [39] in 2015 proposed

a UKF algorithm with a non-linear time series prediction model to build a novel method for RUL's prediction. The UKF and short-term capacity are used to update the battery model state frequently. The proposed model validates higher accuracy and reliability than the extended Kalman filter (EKF) and achieves a mean absolute percentage error (MAPE) and an RMSE of 0.1611 and 0.01156, respectively, at 100 cycles. Li et al. [40] in 2016 proposed a spherical cubature particle filter (SCPF)-based state-space model to examine RUL for 26 Li-ion batteries. The proposed model excels the PF method in terms of prediction accuracy. However, different currents and temperatures could easily affect the model's accuracy. Moreover, the authors [40] proposed a novel prediction method based on the Gaussian process mixture (GPM) for Li-ion batteries' RUL prediction. The proposed model provides more reliable and accurate results than SVM and Gaussian process regression (GPR), achieving RMSEs of 0.0158 and 0.0130 at 60 and 80 inspection cycles. Shen et al. [41] in 2020 introduced a novel online method for RUL prediction of Li-ion batteries with variable discharge current, using a novel two-stage Wiener process model. The authors examined the novel method for 2 Li-ion batteries; achieved a mean absolute error (MAE), RMSE, MAPE, and decision coefficient (R^2), of 3, 3.889, 0.0829, and 0.9751, respectively, for the first battery, and 2.444, 3.122, 0.126, and 0.9856, analogously for the second battery.

Figure 2.8 shows that the Li-ion batteries RUL estimation methods can be mainly categorized into three branches: model-based, data-driven, and stochastic methods. The figure summaries the most important ones mentioned in the literature review.

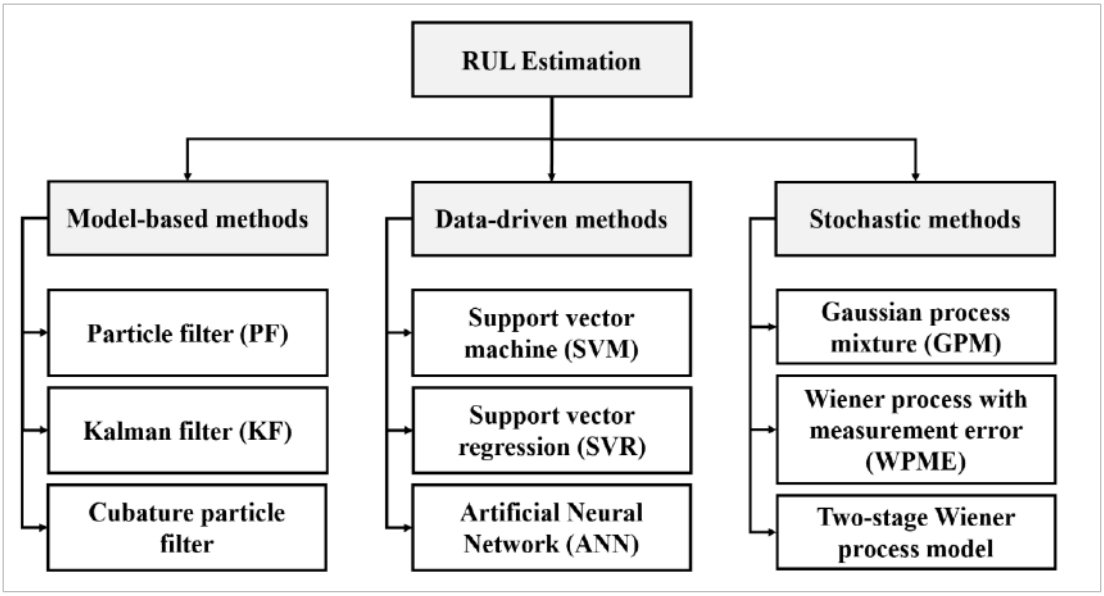


Figure 2.8: Types of RUL estimation methods

It was noted in the previous studies reviewed on the main topic of the thesis, RUL, that the principle used in what was data-driven methods was the classic Nyquist-based data sensing principle, which is time-invariant that does not perform well in the case of non-uniform signals, such as battery cell parameters. Whereas the novel approach proposed by this thesis, the event-driven sensing approach, is time-variant, which acquires data and changes the sampling frequency according to changes in the incoming signal in a manner that the acquisition of the cell parameters is no longer recorded in a preset step of time; instead, it is acquired on the event basis.

CHAPTER 3
RESEARCH METHODOLOGY

This thesis chapter discusses the research methodology in detail, including the system's block diagram, the dataset description, and battery parameters sensing, including classical sensing and event-driven sensing, followed by the feature extraction. Then, it explains the machine learning-based remaining useful life (RUL) prediction using five algorithms. Finally, it shows the models' evaluation techniques, namely percentage split (PS) and cross-validation (CV), and demonstrates the performance evaluation metrics in terms of compression ratio (R_{Comp}) and prediction accuracy.

3.1. System's Block Diagram

Figure 3.1 illustrates the system's block diagram of this work. Firstly, a suitable current high-power Li-ion battery real dataset will be distinguished and utilized during offline processing. Secondly, this dataset will be processed and reconstructed to make its form in line with MATLAB. Using MATLAB, the proposed system will be modeled. The event-driven peak sensing model will efficiently extract the features from the studied battery consumption parameters. It is going to be built primarily based on the phenomenon of shape context feature extraction. The feature extractor will be sufficiently achieved via embedding the event-driven peak sensing phenomenon in the system. These features are going to be used for training the proposed event-driven RUL predictor.

During online processing, the event-driven sensing-based acquired parameters will be used to predict the battery capacity. Besides, by applying an event-driven peak sensing approach, the shape context features will be extracted. Beyond, those fused features will be used for online RUL prediction of the studied Li-ion battery cell using Weka software. The devised solution will excel the counterparts in power consumption due to the event-driven data acquisition and processing approach. The proposed solution is parametrizable and can be utilized in different expected applications such as smart and microgrids, hybrid electric vehicles (HEVs), drones, distributed sensors, and satellites.

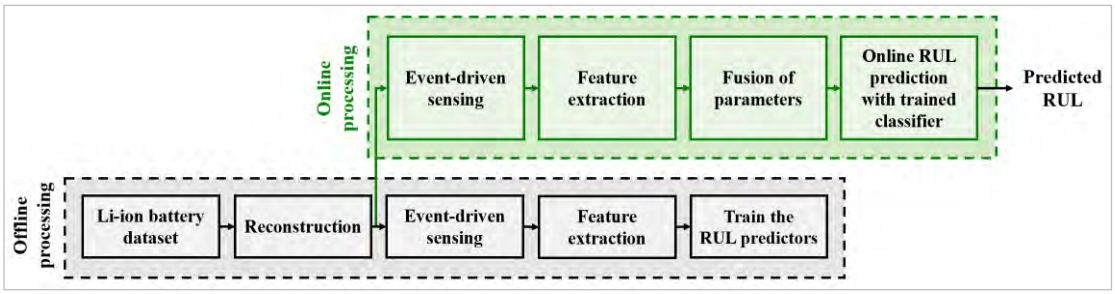


Figure 3.1: The system's block diagram

Certain essential parts should be explained before starting to apply the methodology, as following.

3.2. The Dataset

This work considers the Battery Dataset publicly provided by the NASA Ames Prognostics Center of Excellence (PCoE), a reliable and one of the most authenticated efficient datasets. It is a set of 36 Li-ion cells: #5, 6, 7, 18, and 25 till 56 [42]. To narrow the scope of a research limitation is being processed, this work considers the first cell in this data, cell-5. However, the work can be extended by applying the proposed solution to any of the cells of this dataset. The intended dataset comprises three different operational profiles of the intended high-power Li-ion batteries: charge, discharge, and electrochemical impedance spectroscopy at different temperatures. These batteries were discharged at different load current levels until the voltage dropped to preset thresholds, which the manufacturer recommended to be below 2.7 V to have the effects of deep discharge aging. These batteries were frequently charged and discharged to speed up the aging process. The experiments were only stopped when the intended batteries reached the end of life (EoL) criteria of 30% fade in rated capacity (2 – 1.4 Ahr) [43]. Figure 3.2 shows cell-5 consumption parameters (voltage, current, and temperature) curves for charging/discharging cycle number 31.

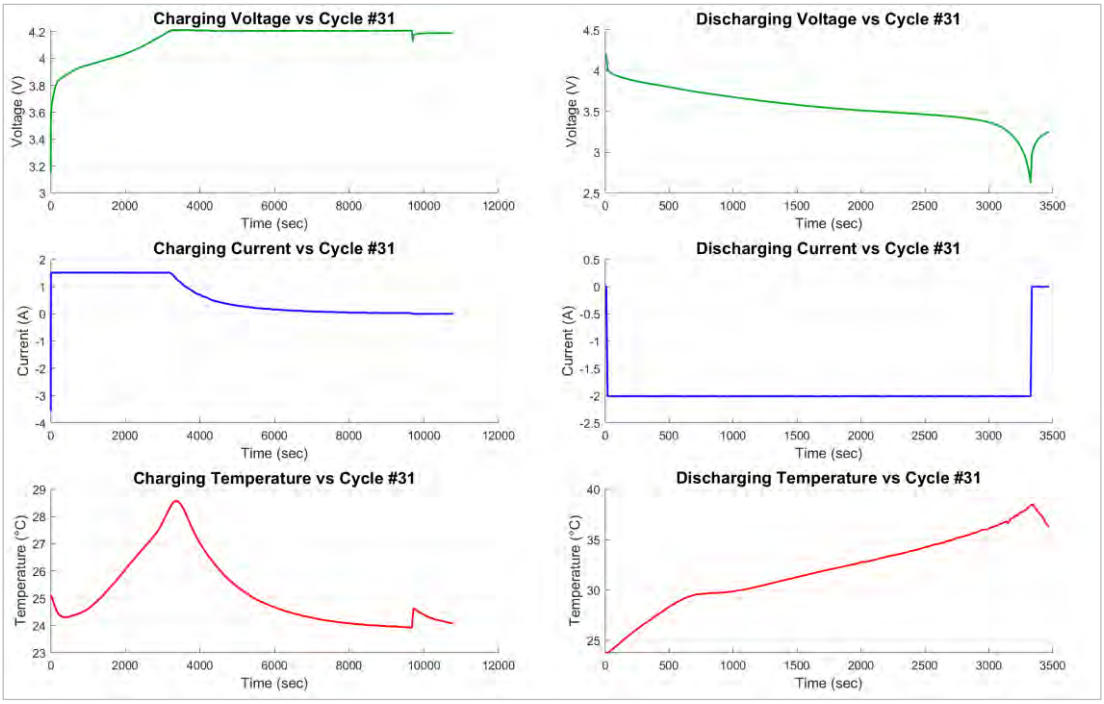


Figure 3.2: Cell-5 consumption parameters curves for charging/discharging cycle number 31

The dataset can be employed in a variety of applications. For example, it can be employed to develop prognostic algorithms since it is data of a large number of run-to-failure time series. Due to differences in depth of discharge (DoD), duration of rest periods, and intrinsic change, no two cells have the same state of life (SoL) in the same period. That prognostic algorithms aim to achieve this uncertainty management indicating the actual use and make accurate RUL predictions in both the end of discharge (EoD) and end of life (EoL) contexts.

3.3. Battery Parameters Sensing

In this thesis work, to prevent processing a large amount of unnecessary data during the sensing process, classical sensing that is constant over time is avoided and replaced by event-driven sensing (EDS) to extract battery cell parameters such as voltages and currents. Therefore, real-time data compression is achieved.

3.3.1. Classical Sensing

As shown in Figure 3.3, there are two types of signals depending on the nature of information: analog and digital signals. An analog signal is a continuous wave with an infinite number of potential values varying within that range. It is represented by a sinusoidal wave. It can be used to measure some physical phenomena changes, such

as light, sound, and temperature. Often, by using analog signals, video and audio communications are transferred.

On the other hand, a digital signal is a discrete wave that carries information in binary form. It has a finite range of possible values, where each sampling point has a discrete value. The more samples, the more accurate representation we get from the input information. A computer easily represents a digital signal. A stream of bits that are either valued by 1 or 0 can be defined for each sample. The digital signal surpasses the analog one. It is more robust against noise; thus, it is clearer and more accurate, and it can be recovered, corrected, and amplified easily. That is why today's trend is to convert the analog signal into digital data, noting that the analog signal is still the best for audio transmissions [44].

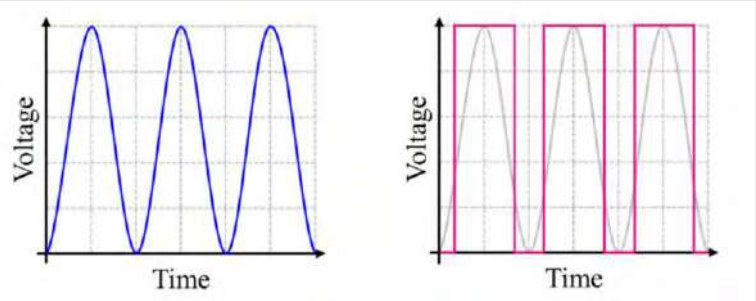


Figure 3.3: Analog signal and digital signal, respectively

There are two methods to achieve analog signal digitization: pulse code modulation (PCM) and delta modulation (DM). PCM is the most common method, in which three processes are applied, as shown in Figure 3.4: analog signal sampling, sampled signal quantizing, and encoding the quantized signal in the form of bitstreams [44].

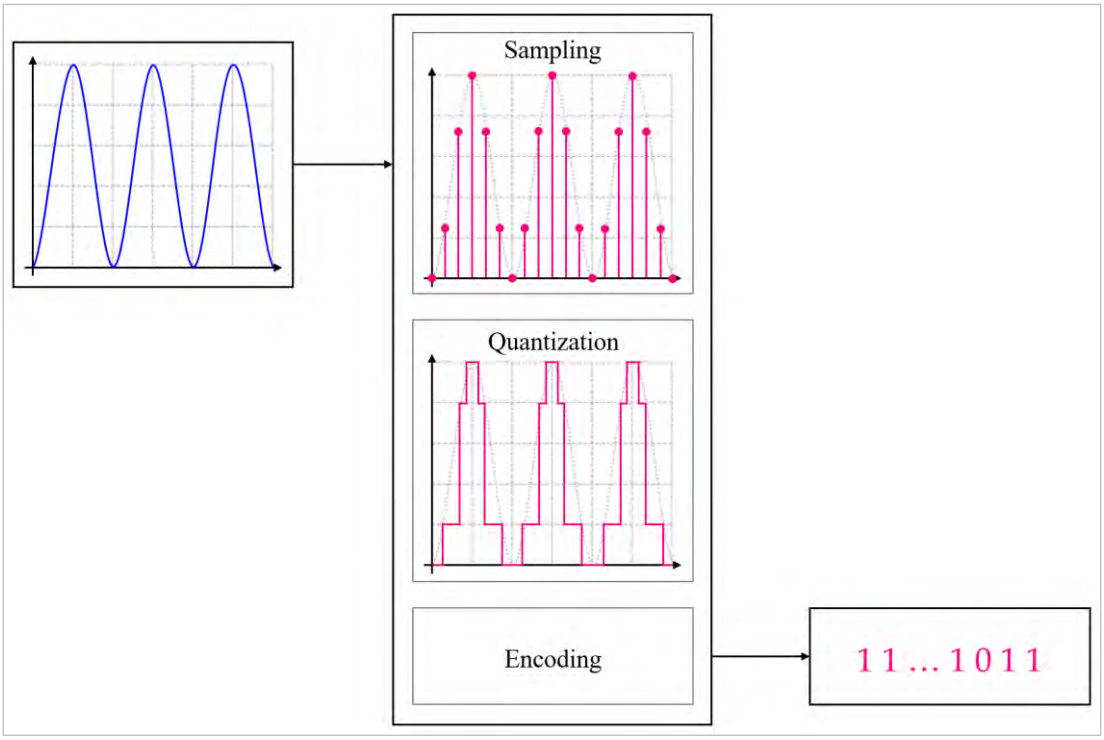


Figure 3.4: PCM processes

Sampling or pulse amplitude modulation (PAM) (as it is referred to, sometimes) is the first analog-to-digital conversion process. However, it should be noted that the signal must be filtered first before sampling to reduce its maximum frequency, as it affects the signal's shape. In the sampling process, the analog signal is sampled every T_s seconds (sampling intervals) to define the waveform shape. As stated before, the more samples we take, the more accurately we will define the waveform [44].

As shown in Figure 3.5 and Equation (2), to get the ideal sample version of an analog (continuous-time) signal, $x(t)$, we are multiplying it with the sampling function, $s(t)$, so we get $x_s(t)$, where $s(t)$ is a train of impulses (δ), as shown in Equation (3).

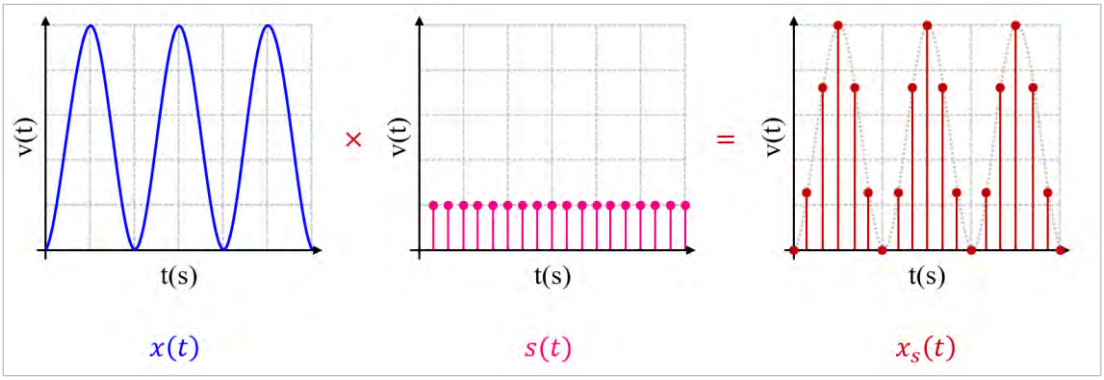


Figure 3.5: Ideal sampling processes

The amplitude of each instant will be defined by the magnitude of $x(t)$ in a discrete version.

$$x_s(t) = x(t) \times s(t) \tag{2}$$

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{3}$$

That is the theoretical ideal sampling function. In practice, it can be defined as shown in Equation (4) to start from 0 since we cannot go back in time until $N - 1$, so a finite number of samples is there.

$$s(t) = \sum_{n=0}^{N-1} \delta(t - nT_s) \tag{4}$$

As shown in Figure 3.6, there are three sampling methods. First, the ideal sampling, where we have a pulse at each T_s. Second, the natural sampling, where we have pulses of short width with various amplitudes. Lastly, the most common sampling method is flat-top sampling, where we have a short width pulse with a single flat-top amplitude [44].

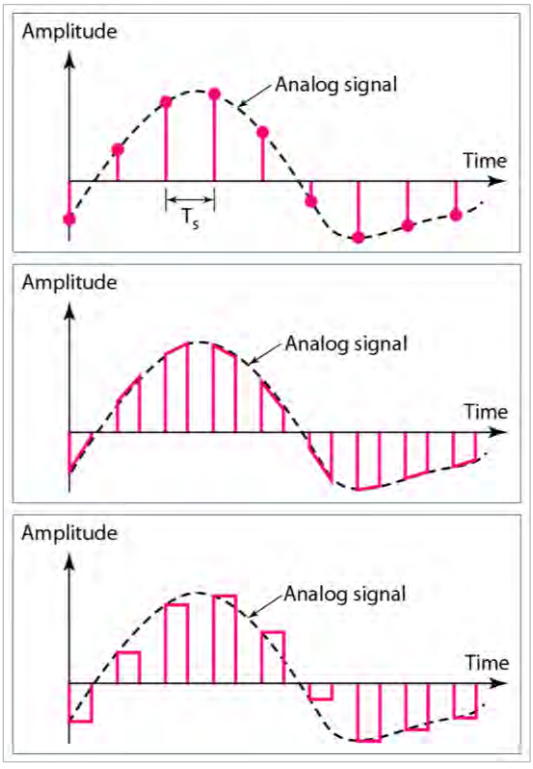


Figure 3.6: Ideal, natural, and flat-top sampling, respectively [44]

In the sampling process, the analog signal is converted into a series of pulses. Each pulse represents the signal's amplitude at a given point. Thus, the result will be a series of samples that keep the analog signal shape. However, it should be noted here that the resulting signal cannot be used in the encoding process since it is still an analog signal [44].

After sampling, the second stage in analog to digital conversion comes, which is the quantification process. In this process, the entire space is partitioned into a non-linear map in which a single value represents each subspace. In this stage, the sampled continuous-valued signals are decomposed into discrete value data. Here it should be noted that quantization is a non-reversible map. Hence, once this process is applied, information is lost, and errors appear. Therefore, reducing the quantization error is one of the most critical issues in this process. Quantization errors can be reduced by increasing the number of units of quantities; however, this will pose a problem in implementation complexity [44].

As shown in Equation (5), to get the quantized version, $x_Q(t)$, of a continuous-time signal, $x(t)$, we are adding the sample function, $x_s(t)$, to the quantization error, $Q_e(t)$, which is a sample signal that is going to be added to each sample value.

$$x_Q(t) = x_s(t) + Q_e(t) \tag{5}$$

As shown in Equation (6) and Figure 3.7, after quantization, the signal will be no longer a continuous-time signal; instead, it will be discrete, $x_Q[n]$, at a given n. Here we can realize that sampling is a discretization in time while quantization is a discretization in magnitude.

$$x_Q[n] = x(t - nT_s) + Q_e(t - nT_s) \tag{6}$$

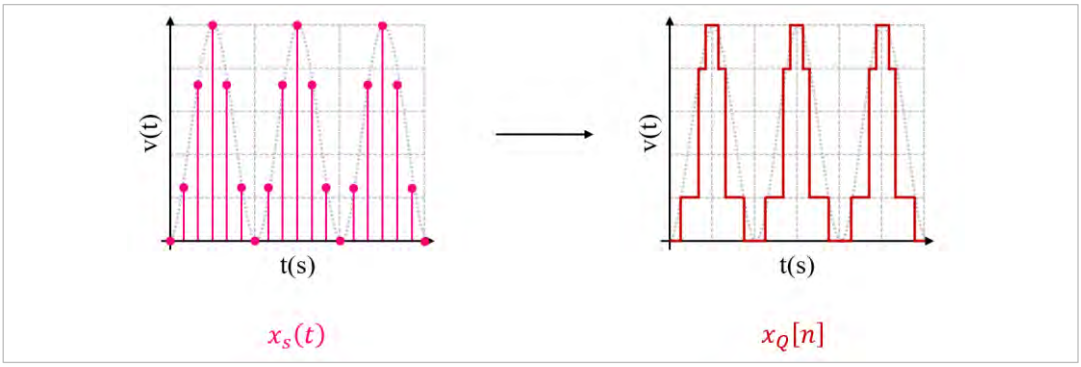


Figure 3.7: Quantization process

The quantization error (see Equation (7)) will define the system's signal-to-noise ratio (SNR), which is expressed in decibels (dB). Its standard formula is shown in Equation (8).

$$Q_e(t - nT_s) = x_Q[n] - x(t - nT_s) \tag{7}$$

$$SNR_{dB} = \frac{P_{Signal}}{P_{Noise}} \approx (6.02 \cdot N + 1.761)dB \tag{8}$$

Where, N is the number of bits per sample in the convertor, defining the convertor's resolution, P_{Signal} is the average signal power, and P_{Noise} is the average noise power. If we keep increasing N, the quantization error, Q_e, will be reduced; hence, SNR will be increased [44].

As shown in Figure 3.8, quantization can be classified into two main classifications: uniform quantization and non-uniform quantization. The quantization step sizes are equally distributed in uniform quantization and vice versa for the non-uniform quantization. The most popular algorithms of non-uniform quantization are the Lloyd max quantization and the μ law quantization [45].

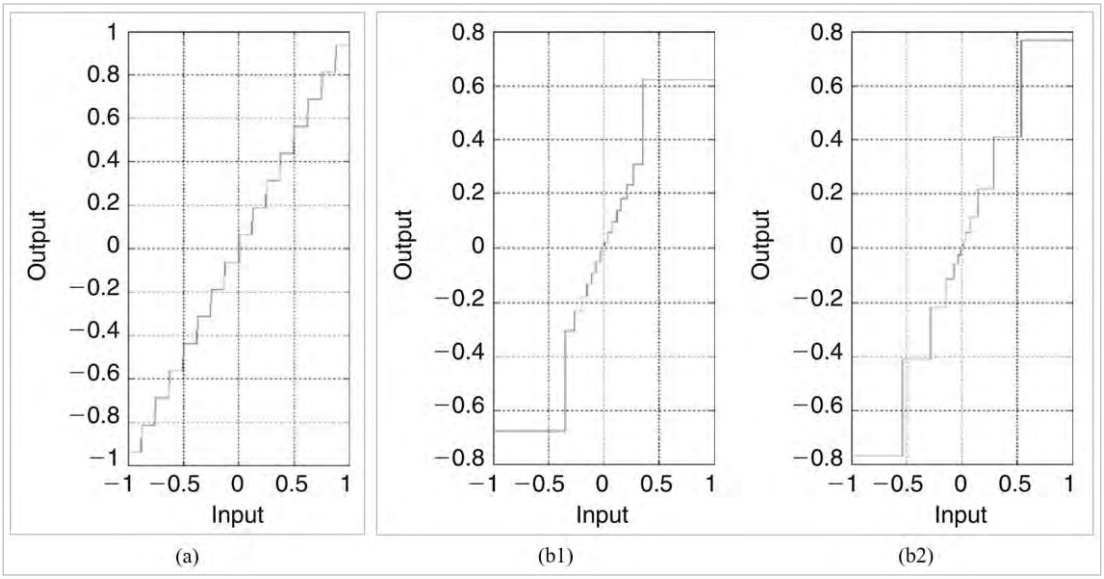


Figure 3.8: (a) Uniform quantization and non-uniform quantization: (b1) Lloyd max quantization and (b2) μ law quantization [45]

As shown in Figure 3.9, the final step in the analog-to-digital conversion process is the encoding, where we assign a unique digital code to each value/state in the

quantized signal, $x_Q[n]$, either by applying 2-bit coding, 3-bit coding, or so on. The more bits, the better representation of the incoming signal we get [44].

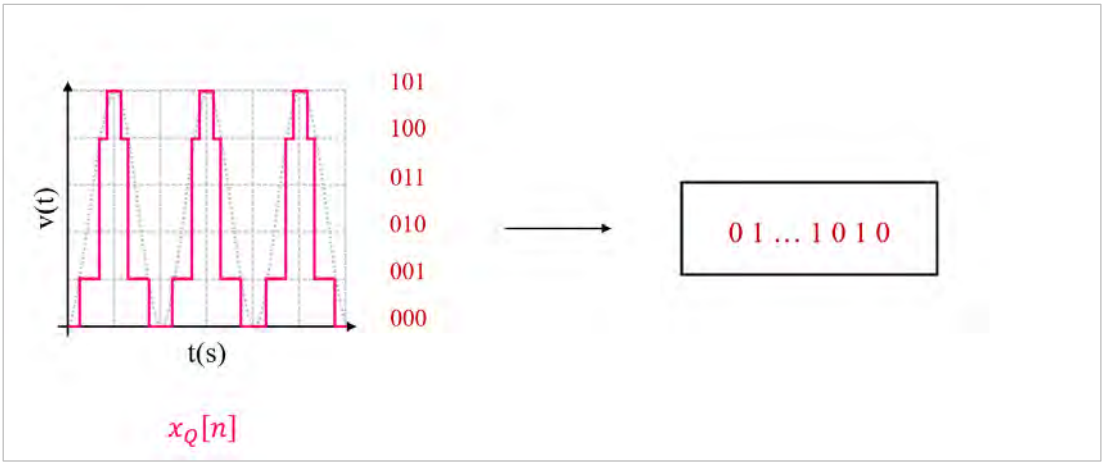


Figure 3.9: Encoding process

3.3.2. Reconstruction and Event-Driven Sensing (EDS)

One of the essential components in the BMSs is the analog-to-digital converter (ADC), from which the classical type takes samples based on Nyquist's theory; thus, it does not show good performance in the case of non-uniform signals, such as battery cell parameters (voltages and currents, for example) [46]–[49]. In contrast, event-driven ADC (EDADC) is an excellent alternative in the said case, as it changes the sampling frequency according to changes in the incoming signal. The main concept of the Event-Driven (ED) approach is that the acquisition of the cell parameters is no longer recorded in a preset step of time but rather on the event basis, using an ADC that is integrated into classical BMS [50].

Figure 3.10 shows the architecture of a successive approximation conversion-based EDADC, and Algorithm 1 shows its working principle. The comparator is comparing the digital value of the incoming analog signal, $\tilde{x}_{n-1}(t)$, with the digital reference signal, $x_{ref}(t)$. Then, it compares the output with the $x_{ref}(t)$, if the digital value of $\tilde{x}_{n-1}(t)$ is greater than $x_{ref}(t)$, the increment is applied by 1, if lower, the decrement is applied by 1; otherwise, nothing occurs, where the $x_{ref}(t) = \frac{1}{2}C$, C is the least significant bit (LSB) or the comparator step, given by Equation (9).

$$C = \frac{\Delta\tilde{x}_{n-1}}{2^N - 1} \tag{9}$$

Where, N is the number of bits per sample in the convertor, defining the convertor's resolution.

Then based on the above, changing the increment and decrement depends on the reference signal comparison, and the number of levels is changed log (number of levels) bits. The new output of $x_n(t)$ is generated corresponding to the previous signal, $\tilde{x}_{n-1}(t)$. Simultaneously, the $x_n(t)$ will be a feedback input through the DAC, which will generate an analog signal based on it to be considered the new reference signal. Then the whole process will repeat.

The information transfer is controlled by bi-directional control signaling, the request, and the acknowledgment. When data is computed in one stage, it sends a request to the next stage; it also sends an acknowledgment when it is ready to receive another data. The total delay of the conversion loop is defined as the time elapsed between the input signal crossing a level and the instant when the correction on the reference signal becomes accurate. When a sample conversion is triggered, the input signal must not cross any quantization level before the ADC can handle new data [51].

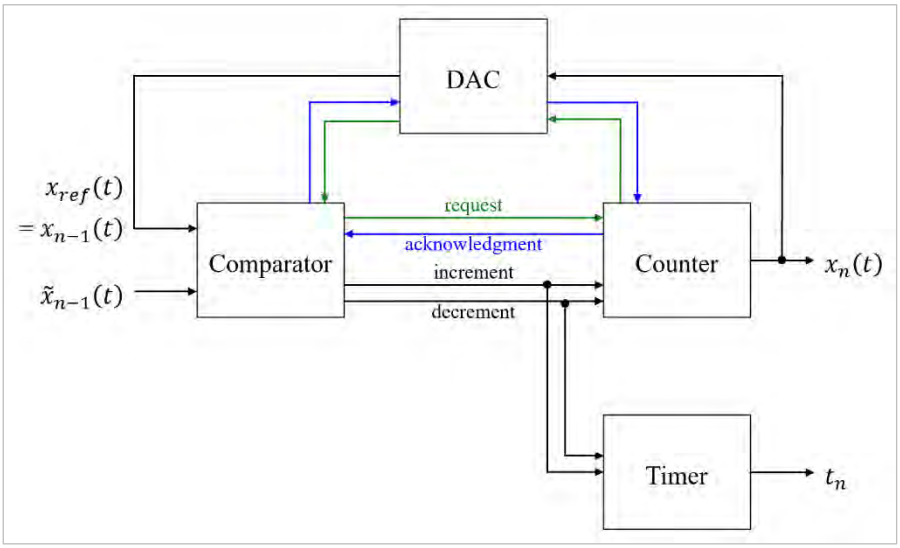


Figure 3.10: EDADC architecture

Algorithm 1

Begin

Initialize all parameters;

$\tilde{x}_{n-1}(t)$ // digital value of the incoming analog signal

$x_{ref}(t)$ // digital reference signal

Compare ($\tilde{x}_{n-1}(t), x_{ref}(t)$)

if $\tilde{x}_{n-1}(t) > x_{ref}(t)$

inc = +1 // the increment is applied by 1

else if $\tilde{x}_{n-1}(t) < x_{ref}(t)$

dec = +1 // decrement is applied by 1

else $\tilde{x}_{n-1}(t) = x_{ref}(t)$

inc = dec = 0 // otherwise nothing occurred

end if

From inc and dec, **generate** $x_n(t)$ // $x_n(t) = Log(N)$; N is the number of levels or bits

While inc != dec != 0 **do** // $x_n(t) = x_{ref}(t)$

Repeat the whole process

Output

$x_n(t)$ // getting by the digital value sample

$\boldsymbol{t_n = t_k - t_{k-1}}$ // the time elapsed since the previous converted sample (given by the timer)

The intended Li-ion battery parameters waveforms are up-sampled by a factor of 100 to assess the event-driven sensing (EDS) element. By following a good mix of anti-aliasing filters and cubic-spline interpolators, up-sampling is accomplished [52]. It offers a quasi-analog representation of the battery parameters waveforms, which can be used as input of the EDS component.

In the case of EDADC, the sampling frequency is adjusted according to changes in the analog signals, $\tilde{x}(t)$. Samples are acquired as a function of the preset thresholds [48], [49]. The approach is given by Equation (10) and shown in Figure 3.11.

$$t_n = t_{n-1} + dt_n \tag{10}$$

Where, t_n is the present sampling instant which is dt_n instants later than the previous one, t_{n-1}, as shown in Figure 3.11. Likewise, x_n is the present sampling amplitude, and x_{n-1} is the previous one. The pair (x_n, t_n) is equivalent to one of the preset EDS thresholds.

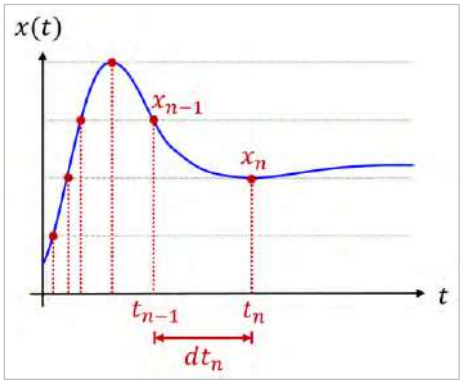


Figure 3.11: Event-driven sensing (EDS)

The EDADC only acquires the relevant data and ignores the rest. As a result, the acquired samples have a noticeable real-time reduction and compression gain compared to their classical counterparts. Furthermore, the EDADC reduces pre-processing activity and improves system processing and power consumption efficiency [46]–[49], [53].

3.4. Feature Extraction

Battery cell charging/discharging curves for voltages and currents indicate valuable information about battery life. Besides, when working individually on the charging and discharging cycles of the intended cell, the shape-dependent on the cell charging/discharging voltages and currents can be extracted. By fusing these features, the current state of the intended Li-ion battery cell can be determined, allowing the stage of determining its RUL to begin.

In this work, an event-driven approach with multi thresholds is applied to extract features once there is an intersection. These features are extracted based on the shape context. While individually considering the intended cell's charging and discharging cycles, it is possible to extract the shape-depended content of the cell charging and discharging voltages, currents [28], and temperature. Onward, these features fused can be used for determining the current state of the considered Li-ion cell.

For example, while considering the charging cycles, the cell terminal voltage features for the k^{th} charging cycle can be extracted using Equation (11).

$$\left(t_{ED_k}, TH_b\right), where\ V_{C_k} = TH_b, and\ k = 1, 2, \dots, M \quad (11)$$

Where, t_{ED_k} is the k^{th} instant when the battery terminal voltage reaches the preset threshold, TH_b, b is the threshold index, V_{C_k} is the k^{th} cell terminal voltage while considering the charging cycles, and M is the total number of cycles that have been considered.

The same approach is taken for the discharging cycles; the cell terminal voltage features for the k^{th} discharging cycle can be extracted using Equation (12).

$$\left(t_{ED_k}, TH_b\right), where\ V_{D_k} = TH_b, and\ k = 1, 2, \dots, M \quad (12)$$

Where, V_{D_k} is the k^{th} cell terminal voltage while considering the discharging cycles.

The intended battery cell features of each charging/discharging cycle voltage, current, and temperature waveform are extracted based on the same principle. This critical step represents the core of the scientific contribution of this work. To achieve it, a MATLAB code has been developed to model the proposed system.

3.5. Machine Learning-Based Remaining Useful Life (RUL) Predictors

This thesis uses machine learning-based RUL predictors. The focus is on k-Nearest Neighbor (kNN)-, Artificial Neural Network (ANN)-, Linear Regression (LR)-, Random Forest (RF)-, and Random Tree (RT)-based predictors. The mentioned predictors have long been considered the most commonly used machine learning (ML) paradigms and have proven capable of learning and predicting more complex data patterns in many application systems. Weka 3.8.5 software, which is open-source software written in Java language, is employed to obtain the prediction results of the intended cell capacity for each of the five machine learning (ML) algorithms used (kNN, ANN, LR, RT, and RF), using two main evaluation techniques: percentage split (PS) and cross-validation (CV). In the PS technique, ratios of 50:50%, 60:40%, 70:30%, 80:20%, and 90:10% are applied to split the data for training and testing purposes. The CV technique is applied for 3-fold, 5-fold, and 10-fold.

3.5.1. K-Nearest Neighbor (kNN)

The k-Nearest Neighbor (kNN) is a similarity-based algorithm that works in a fast and efficient manner. It constructs a non-linear kernel regression model and predicts the regression of a given test element's response by averaging the responses of k's closest neighbors to that element [54]. As shown in Figure 3.12, we have classes '+' and '−'; the algorithm aims to find the class for the element '×.' In this case, we have five nearest neighbors (k=5) which are near to '×': three of them are in the '+' class, and two are in the '−' class. Thus, the element '×' class will be '+.' That is how the kNN algorithm works. The kNN algorithm classifies the testing set by comparing it to the k-nearest neighbors from the training dataset. In this work, by running *weka.classifiers.lazy.IBk* for the kNN model configurations tuning in Weka, it is found that the best parameters that give the highest performance are setting k=3 and choosing the Euclidean distance metric for finding the nearest neighbors. The Euclidean distance metric work principle is shown in Figure 3.13.

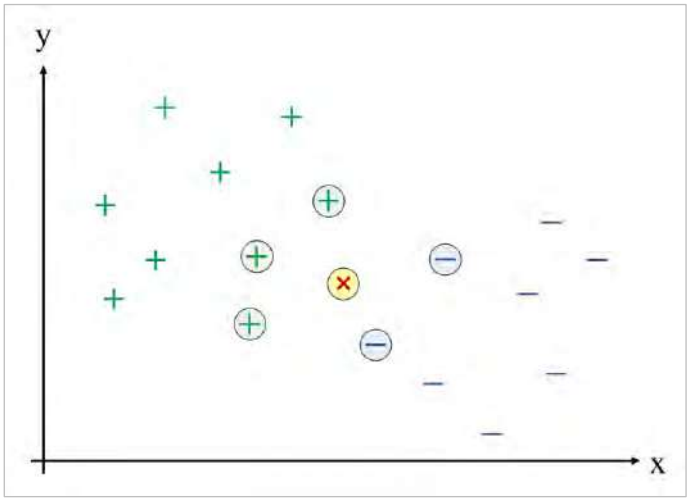


Figure 3.12: K-Nearest Neighbor (kNN) example

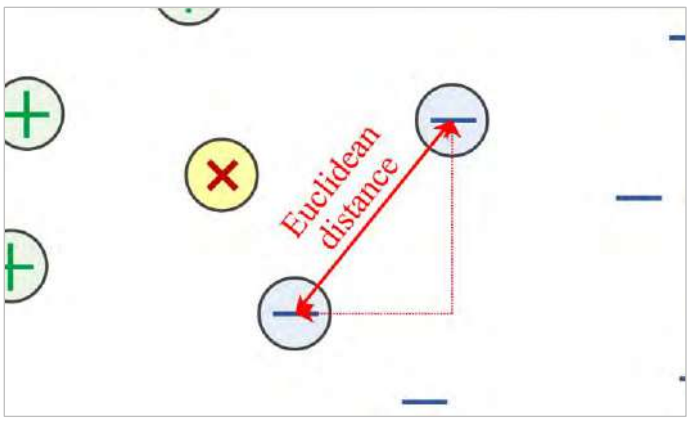


Figure 3.13: Euclidean distance metric example

3.5.2. Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is one of the most robust and efficient algorithms for predicting complex data, which linear classification algorithms cannot succeed. ANNs simulate the biological brain, and the general architecture of any of them consists of neurons representing processing units on three layers, as shown in Figure 3.14, where the hidden layer is employed to proceed with the non-linear classification. Adjusting the harmony between those three layers determines overall network performance. The more hidden layers, the lower the error rate [54]. In this work, by running *weka.classifiers.functions.MultilayerPerceptron* for the ANN model configurations tuning in Weka, it is found that the best parameters that give the highest performance are setting the number of hidden layers to 5 and the number of iterations to 1000. The gradient descent momentum is used with a variable learning rate till a satisfying prediction rate is reached. The Radial Basis function is set for neuron activation.

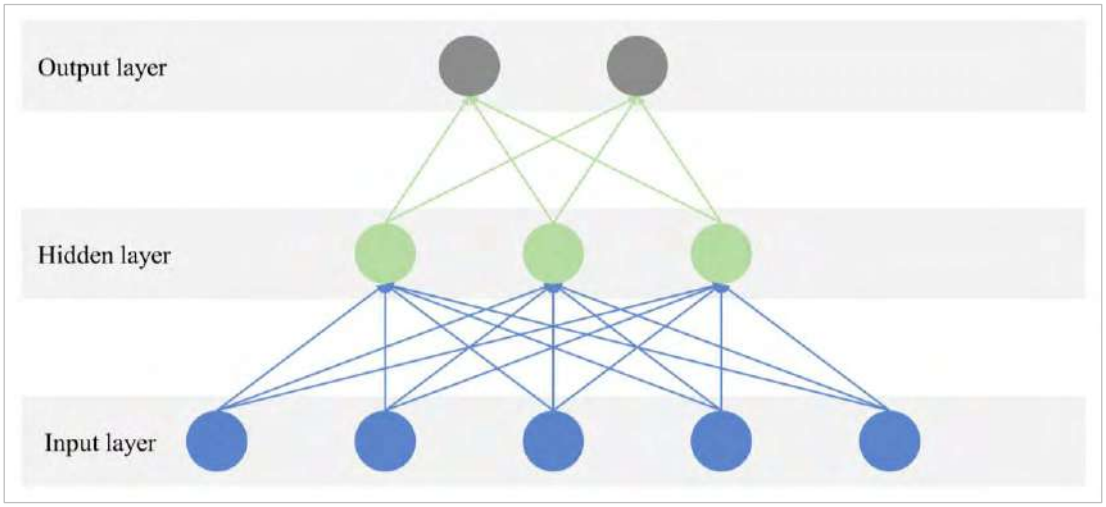


Figure 3.14: Artificial neural network (ANN) model

3.5.3. Linear Regression (LR)

Linear Regression (LR) is the most straightforward ML algorithm that can work on complex data patterns. It was built on the premise that the relationship between the input and output variables under study is linear. Thus, this algorithm employs the statistical model that predicts the relationship between the variables based on a linear equation [55]. Figure 3.15 shows an example of the LR algorithm model. This work uses the least square linear regression algorithm by running *weka.classifiers.functions.LinearRegression* for the LR model in Weka. The intercept is

set to true and is used in model calculations. It is found that the best parameter that gives the highest performance is setting the value of the complexity parameter controls the amount of shrinking to 0.5.

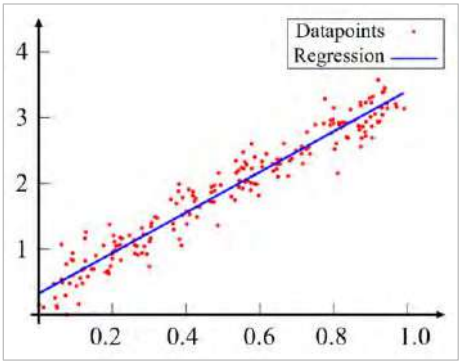


Figure 3.15: Linear Regression (LR) model

3.5.4. Random Tree (RT)

The Random Tree (RT), or the decision tree, is one of the best ensemble supervised ML algorithms [56]. Its working principle is that a robust learner can be built by uniting many weak, randomly formed learners. It creates many decision branches at different nodes; each terminal leaf produces a random classification of the data sample input; being random indicates that all branches obtained equal probabilities of being sampled. The RT collects the produced decisions and gives a final output with the prediction that got the majority votes [57]. Figure 3.16 shows a schematic diagram of the RT algorithm. In this work, by running *weka.classifiers.trees.RandomTree* for the RT model configurations tuning in Weka, it is found that the best parameter that gives the highest performance is setting the depth of the tree to 15. The algorithm randomly selects the split based on the information gain criterion. For each split, the entropy of each child node is calculated. Onward, the entropy of the split is computed as the weighted average entropy of child nodes. The splitting criteria used to select a split is the highest information gain splitting.

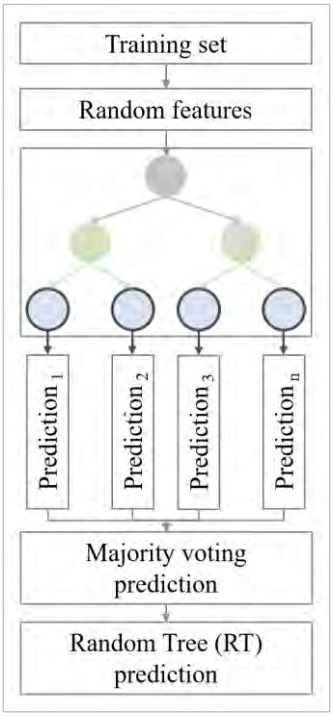


Figure 3.16: Random Tree (RT) algorithm schematic diagram

3.5.5. Random Forest (RF)

The Random Forest (RF) algorithm creates a forest of random trees, i.e., creates many random decision trees at different nodes to get a more accurate prediction than a single random tree algorithm. Each tree in the RF produces a random classification of the data sample input; the RF collects these decisions and gives the output which had the highest vote decision [57]. Figure 3.17 shows a schematic diagram of the RF algorithm. In this work, by running *weka.classifiers.trees.RandomForest* for the RF model configurations tuning in Weka, it is found that the best parameters that give the highest performance are setting the number of trees to 50, each with 10 branches. The splitting criteria used is the interaction-curvature.

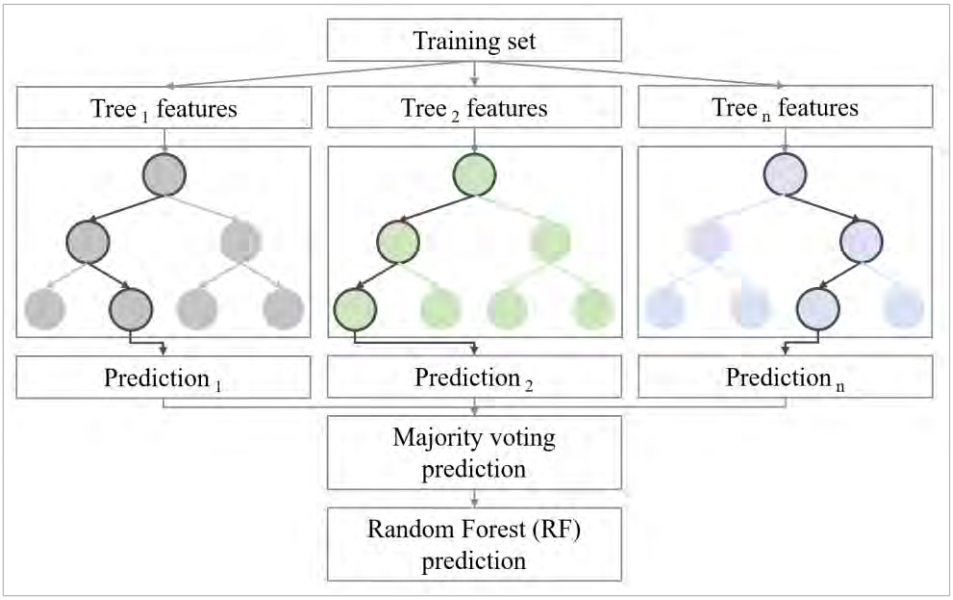


Figure 3.17: Random Forest (RF) algorithm schematic diagram

3.6. Models' Evaluation Techniques

This work uses two techniques to evaluate the models on the predicted data, namely percentage split (PS) and cross-validation (CV).

3.6.1. The Percentage Split (PS)

Percentage split (PS) is a prevalent model-selection technique in ML applications. It is used for the performance evaluation of model prediction algorithms. According to PS, the original dataset is divided into training and testing sub-datasets based on a specific ratio. While the training subset is utilized to fit the ML model on available data, the testing subset is mainly used for evaluating the fit model. Specifically, the performance of the fit ML model is evaluated based on new data not previously utilized in the training phase. That is achieved by comparing the expected values and predictions of the model. In fact, there are several options for the selection of the training-to-testing ratio including 50:50%, 60:40%, 70:30%, 80:20%, and 90:10% as illustrated in Figure 3.18. The optimal ratio can be selected for a specific problem considering computational cost and representativeness of both training and testing sub-datasets. Practically speaking, it is more desirable to select a 2:1 ratio as it delivers acceptable and fair results [58]. However, it is worth mentioning that although PS is fast and simple to employ, it is not suitable for all ML problems. For example, it is suitable for large-enough dataset problems and may not perform well for small or unbalanced

datasets [59]. In such cases, k-fold cross-validation (CV) would be more suitable for model evaluation.

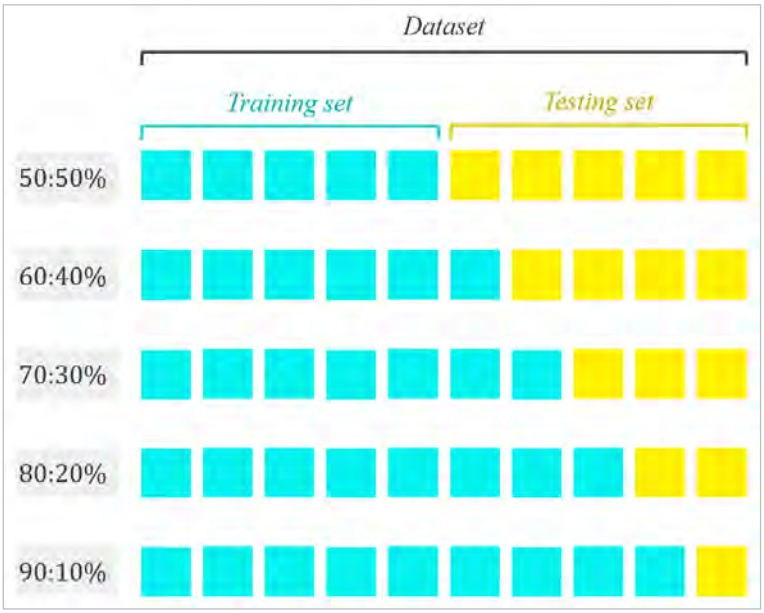


Figure 3.18: Example of 50:50%, 60:40%, 70:30%, 80:20%, and 90:10% training and testing PS

3.6.2. The Cross-Validation (CV)

Cross-validation (CV) is a well-known technique widely employed in ML approaches, such as ANNs, learning-based techniques, and decision trees. It is utilized for various purposes such as accuracy evaluation, tuning parameters, and feature selection. Besides, it is commonly used in building predictive models to evaluate the model's generalization capability such that the overfitting can be avoided. It is also used in the feature selection process of model building. It is worth mentioning that to avoid biased estimation, predictive features should be selected based only on the training set. In selecting these features based on the whole learning set, the testing set's data will be then used for both predictions and validation purposes. That results in information leakage and leads to downwards biased estimation. In other words, it underestimates the actual prediction error. CV might also be utilized for tuning the model parameters, such as the optimum number of the nearest neighbors in the kNN classification approach. In this case, this technique is performed multiple times for various parameter values, and the optimum parameter that reduces the CV error is then selected for constructing the final model. That generally results in reducing the model overfitting [60].

The data is generally divided into two sub-sets in CV, i.e., the training (calibration) and the testing (validation) sets. The training set is of size $(n - v)$, and it is used to approximate the pattern. On the other hand, the testing set is of size v and is used to estimate the predicted variance. The procedure is repeated for all possible sub-division of the dataset with appropriate sizes. The CV principle is then directly obtained as the average of projected predicted differences over these repetitions.

Furthermore, k-fold CV is one of the techniques that are commonly utilized for estimating the classification error. In this technique, the data set is split into k-folds, wherein the training phase of the classifier is conducted using $k - 1$ folds. The error value in the testing phase is then obtained based on the remaining fold. The average value of the errors at each fold represents the k-fold CV error estimation. Figure 3.19 illustrates an example of a 5-fold CV.

Figure 3.19: Example of a 5-fold CV

3.7. Performance Evaluation Metrics

3.7.1. Compression Ratio (R_{Comp})

The compression ratio (R_{Comp}) for the data worked on is given by Equation (13) [53].

$$R_{Comp} = \frac{N_{classic}}{N_{ED}} \tag{13}$$

Where, N denotes the number of samples acquired in the classical counterpart, and N_{ED} denotes the number of samples acquired in the suggested model.

3.7.2. Prediction Accuracy

The accuracy of considered Li-ion battery capacity prediction, hence RUL prediction, is computed in terms of the correlation coefficient (r), the mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE).

The **correlation coefficient (r)** measure shows how strong the relationship between two variable series is, for example, the actual and predicted capacities. It takes a value from -1 to 1. Inverse relationship (perfect anti-correlated) is indicated by -1, and direct relationship (perfect correlated) is indicated by 1, while the uncorrected series (no correlation) is represented by 0 [61]. The correlation coefficient, r, can be calculated automatically using Microsoft Excel's *CORREL* function, or manually as shown in Equation (14) [61].

$$r = \frac{\sum_{i=1}^{n}\left(c_i^{act} - \overline{c^{act}}\right)\left(c_i^{pred} - \overline{c^{pred}}\right)}{\sqrt{\sum_{i=1}^{n}\left(c_i^{act} - \overline{c^{act}}\right)^2 \sum_{i=1}^{n}\left(c_i^{pred} - \overline{c^{pred}}\right)^2}} \tag{14}$$

Where, c_i^{act} is the actual capacity for the i^{th} instant, c_i^{pred} is the predicted capacity, $\overline{c^{act}}$ and $\overline{c^{pred}}$ are the mean of each one analogously, and n is the total number of the instants.

The **mean absolute error (MAE)** and **root mean squared error (RMSE)** are given by Equations (15) and (16), respectively [62].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|c_i^{act} - c_i^{pred}\right| \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(c_i^{act} - c_i^{pred}\right)^2}{n}} \tag{16}$$

Where, n is the data points or response values.

The MAE estimates the average error, while the RMSE estimates the standard deviation of the random component in the data. The closer their values to 0, the smaller the error value, and hence the fit is more beneficial for prediction. Both the MAE and RMSE help compare models in terms of accuracy if their errors are in similar units.

The **relative absolute error (RAE)** and **root relative squared error (RRSE)** are given by Equations (17) and (18), respectively [62].

$$RAE = \frac{\sum_{i=1}^{n}\left|c_i^{act} - c_i^{pred}\right|}{\sum_{i=1}^{n}\left|c_i^{act} - \overline{c^{act}}\right|} \tag{17}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}\left(c_i^{act} - c_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(c_i^{act} - \overline{c^{act}}\right)^2}} \tag{18}$$

Both the RAE and RRSE range from 0 to 1. The lower their values, the better the prediction; the closer their values are to 0, the more useful for prediction is the fit. Their calculation includes division by the c_i^{act} variation; thus, they are called *relative* errors. Hence, they help compare models in terms of accuracy even if their errors are not in similar units, unlike the MAE and RMSE, which are not applicable if the error units are different [62].

CHAPTER 4
RESULTS AND DISCUSSIONS

This thesis chapter presents and discusses the results on Feature Extraction and Data Compression, Capacity Prediction, which can be translated into the RUL, and Prediction Accuracy.

4.1. Feature Extraction and Data Compression

Considering the dataset of cell-5 from the intended dataset, shape context-based features of the cell charging/discharging voltages, currents, and temperatures are acquired for different charging/discharging cycles. In total, 338 cycles are considered: 170 charging cycles and 168 discharging cycles. Each cycle is considered as an instant and presented by corresponding voltage, current, and temperature curves. Cell-5 features for all cycles are extracted using a MATLAB-based event-driven sampling developed model. Table 4.1 shows samples of the considered cell's extracted features values, where k is the cycle number and the extracted feature represented by the pair $(Time, Preset\ threshold)$, which is the first intersection with the preset threshold; the time, voltage, current, and temperature are measured in seconds, volts, amperes, and °C, respectively. The extracted features will represent the attributes for the various machine learning (ML) algorithms that are applied using Weka software, as will be shown in the next section.

Table 4.1: Samples of the cell-5 charging/discharging extracted features
(for cycles number 31, 71, 101, and 152)

k / **Feature**	**31**	**71**	**101**	**152**
	Extracted Feature ($Time, Preset\ threshold$)			
(1a) Charging Voltage	(1598.521, 4.00)	(1101.452, 4.00)	(755.186, 4.00)	(296.587, 4.00)
	(2129.351, 4.05)	(1582.122, 4.05)	(1176.799, 4.05)	(647.083, 4.05)
	(2517.725, 4.10)	(1954.103, 4.10)	(1530.230, 4.10)	(992.264, 4.10)
	(2844.983, 4.15)	(2274.571, 4.15)	(1834.196, 4.15)	(1283.398, 4.15)

Feature \ k	31	71	101	152
	Extracted Feature ($Time, Preset\ threshold$)			
(1b) Discharging Voltage	(3235.924, 3.1)	(2790.606, 3.1)	(2518.737, 3.1)	(2243.792, 3.1)
	(3187.192, 3.2)	(2726.423, 3.2)	(2441.178, 3.2)	(2148.640, 3.2)
	(3106.270, 3.3)	(2617.989, 3.3)	(2307.497, 3.3)	(1970.762, 3.3)
	(2900.130, 3.4)	(2334.647, 3.4)	(1966.204, 3.4)	(1573.212, 3.4)
	(2136.073, 3.5)	(1662.866, 3.5)	(1370.145, 3.5)	(1081.448, 3.5)
	(1390.320, 3.6)	(1091.170, 3.6)	(908.924, 3.6)	(723.139, 3.6)
	(892.620, 3.7)	(690.774, 3.7)	(571.227, 3.7)	(448.639, 3.7)
	(499.600, 3.8)	(383.660, 3.8)	(297.211, 3.8)	(207.675, 3.8)
(1c) Discharging Load Voltage	(3.341, 1.5)	(3.343, 1.5)	(3.333, 1.5)	(3.339, 1.5)
	(4.009, 1.8)	(4.011, 1.8)	(4.000, 1.8)	(4.007, 1.8)
	(4.677, 2.1)	(4.680, 2.1)	(4.666, 2.1)	(4.675, 2.1)
	(5.345, 2.4)	(5.348, 2.4)	(5.333, 2.4)	(5.343, 2.4)
(2a) Charging Current	(4.917, 0.5)	(4.670, 0.5)	(4.612, 0.5)	(4.721, 0.5)
	(5.094, 0.8)	(4.824, 0.8)	(4.760, 0.8)	(4.874, 0.8)
	(5.272, 1.1)	(4.977, 1.1)	(4.907, 1.1)	(5.026, 1.1)
	(5.450, 1.4)	(5.130, 1.4)	(5.055, 1.4)	(5.178, 1.4)
(2b) Discharging Current	(14.459, -1)	(14.423, -1)	(14.415, -1)	(14.436, -1)
(2c) Discharging Load Current	(14.487, 1)	(14.472, 1)	(14.448, 1)	(14.464, 1)
(3a) Charging Temperature	(2208.833, 26.4)	(1435.746, 26.4)	(834.671, 26.4)	(945.836, 26.4)
	(2591.407, 27.0)	(1800.769, 27.0)	(1231.978, 27.0)	(1153.742, 27.0)
	(2929.346, 27.6)	(2136.510, 27.6)	(1551.721, 27.6)	(1344.147, 27.6)
	(3152.501, 28.2)	(2386.867, 28.2)	(1826.388, 28.2)	(1514.680, 28.2)

Feature \ k	31	71	101	152
	Extracted Feature ($Time, Preset\ threshold$)			
(3b) Discharging Temperature	(1407.337, 31)	(1174.062, 31)	(1000.122, 31)	(893.776, 31)
	(1744.043, 32)	(1436.003, 32)	(1222.501, 32)	(1074.096, 32)
	(2107.117, 33)	(1711.312, 33)	(1448.328, 33)	(1263.659, 33)
	(2438.347, 34)	(1985.678, 34)	(1677.052, 34)	(1450.203, 34)
	(2721.665, 35)	(2239.134, 35)	(1900.631, 35)	(1636.540, 35)
	(2949.530, 36)	(2461.343, 36)	(2104.432, 36)	(1810.195, 36)
	(3168.952, 37)	(2649.302, 37)	(2286.368, 37)	(1976.171, 37)
	(3291.785, 38)	(2786.959, 38)	(2443.773, 38)	(2124.757, 38)

The feature extraction principle is clear from Figure 4.1 till Figure 4.3, demonstrating the Table 4.1 corresponding charging/discharging voltage, current, and temperature curves, respectively, for the considered cycles samples with preset thresholds.

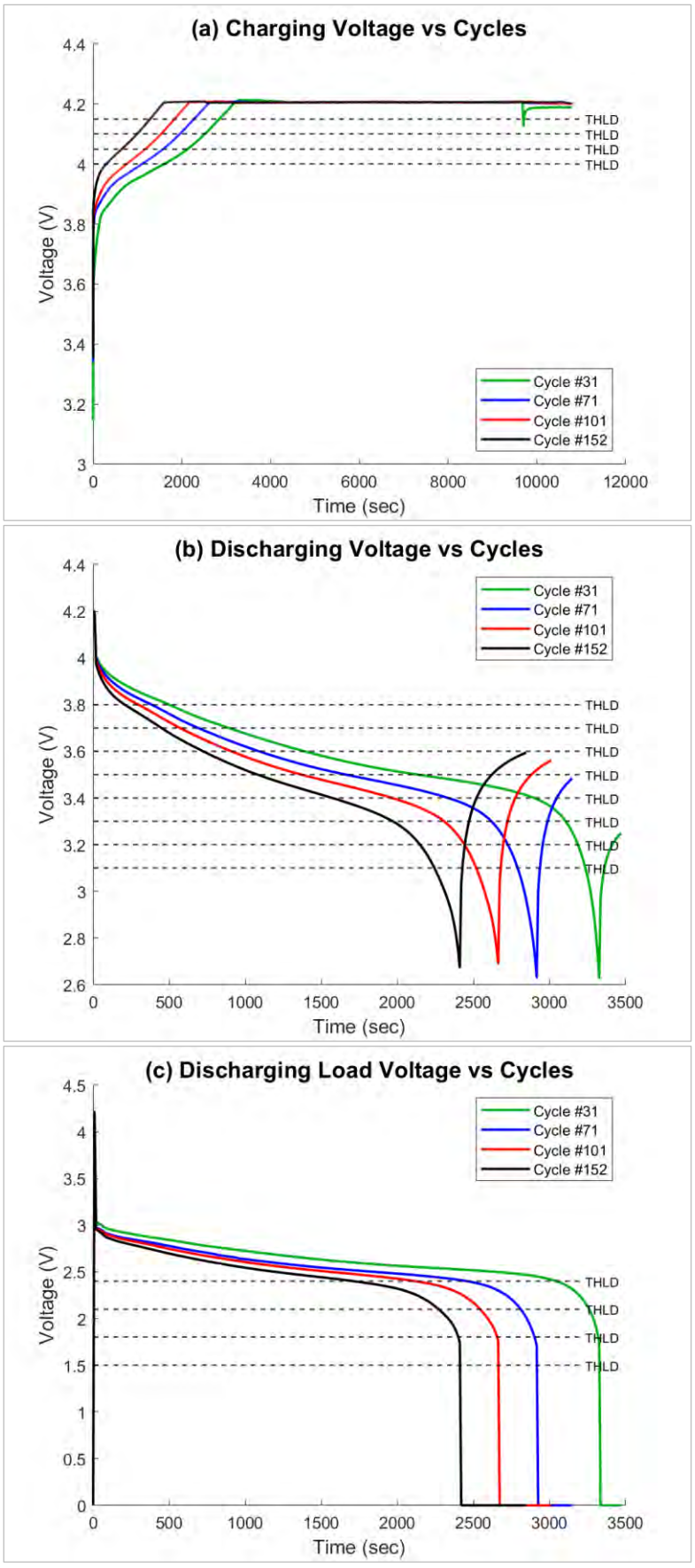


Figure 4.1: Samples of the cell-5 charging/discharging voltage curves (for cycles number 31, 71, 101, and 152) with preset thresholds

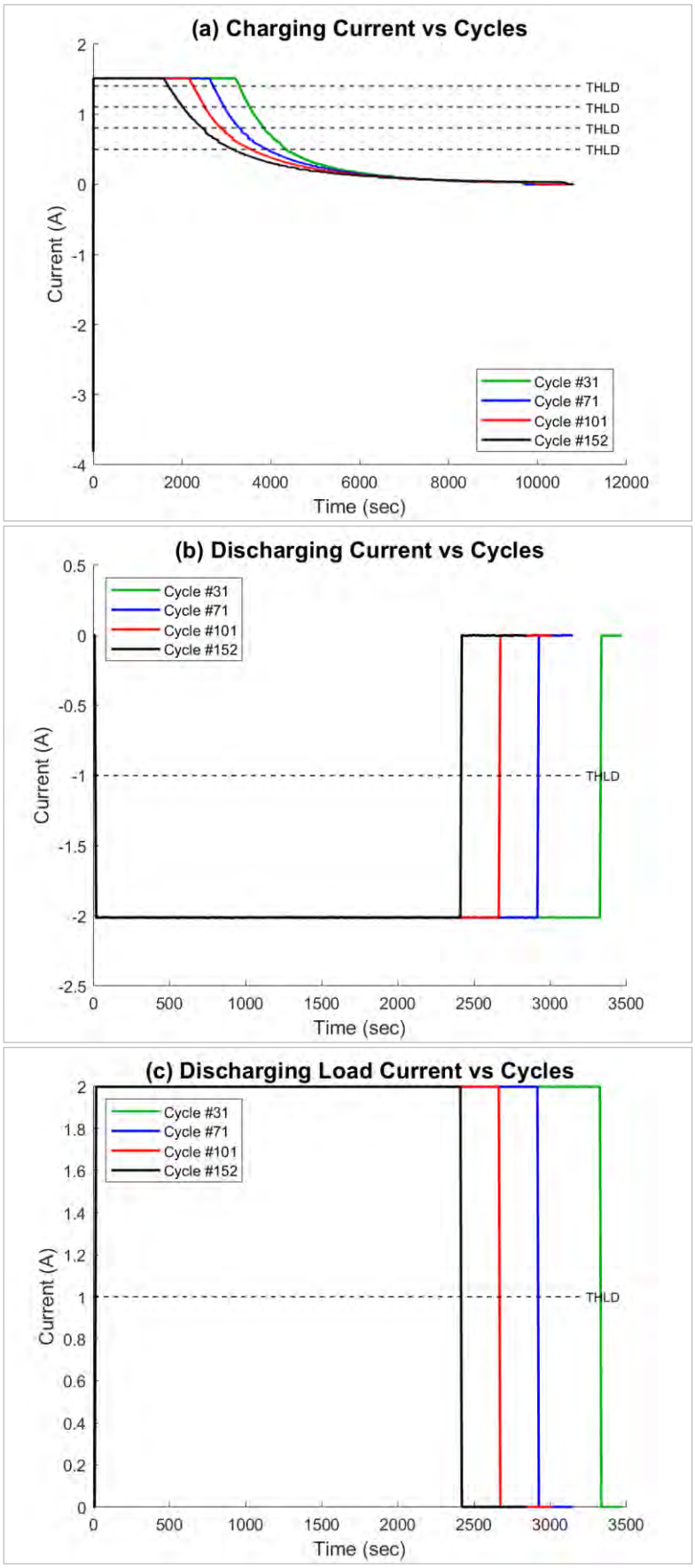


Figure 4.2: Samples of the cell-5 charging/discharging current curves (for cycles number 31, 71, 101, and 152) with preset thresholds

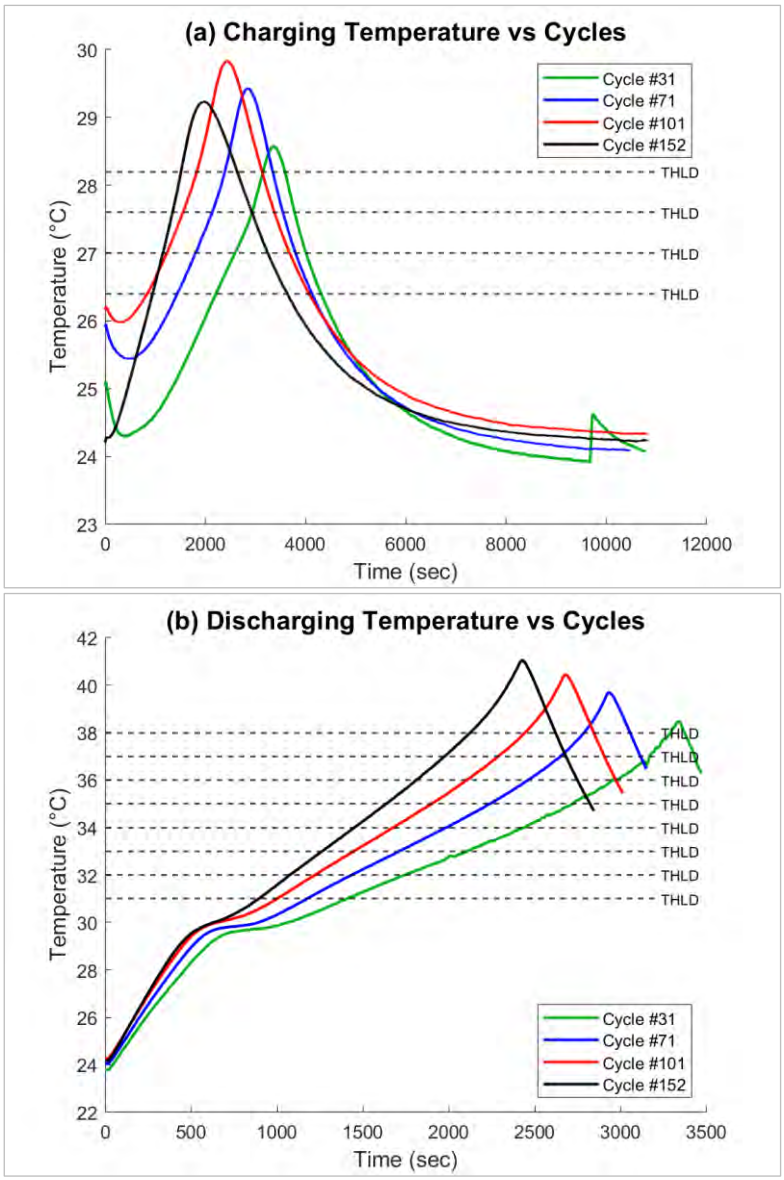


Figure 4.3: Samples of the cell-5 charging/discharging temperature curves (for cycles number 31, 71, 101, and 152) with preset thresholds

As mentioned earlier, traditional sensing causes a lot of unnecessary data to be processed; hence, it impedes real-time data compression. In the studied case and the context of data compression, each instant is composed of three 3,500 seconds long waveforms. Therefore, each instant consists of 10,500 samples while acquiring it at a constant rate of 10 Hz [63]. On the other hand, the suggested event-based solution acquires only 24 features per instant. The compression ratio is calculated using Equation (13), indicating that the suggested solution's average compression gain is 437.5-fold. That undoubtedly leads to a significant improvement in power consumption.

4.2. Capacity Prediction

Considering the pre-mentioned extracted features of cell-5 using the MATLAB-based event-driven sampling developed model, in total, 168 instants are considered. Weka software is employed to obtain the prediction results of the intended cell capacity for each of the five ML algorithms used (kNN, ANN, LR, RT, and RF), using two main evaluation techniques: percentage split (PS) and cross-validation (CV). In the PS technique, ratios of 50:50%, 60:40%, 70:30%, 80:20%, and 90:10% are applied to split the data for training and testing purposes. The CV technique is applied for 3-fold, 5-fold, and 10-fold.

4.2.1. K-Nearest Neighbor (kNN)

Table 4.2 shows the cell-5 actual and predicted capacity for training and testing split of 70:30% using the kNN algorithm and the absolute error values, with a total number of instants of 50.

Table 4.2: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the kNN

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.496	1.501	0.005
2	1.480	1.481	0.000
3	1.339	1.339	0.000
4	1.852	1.842	0.010
5	1.586	1.583	0.002
6	1.318	1.317	0.001
7	1.433	1.432	0.002
8	1.288	1.296	0.009
9	1.814	1.821	0.007
10	1.365	1.369	0.004
11	1.799	1.803	0.003
12	1.355	1.360	0.005
13	1.412	1.411	0.002
14	1.836	1.840	0.004
15	1.560	1.556	0.004
16	1.313	1.314	0.000

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
17	1.417	1.423	0.005
18	1.518	1.524	0.007
19	1.319	1.317	0.002
20	1.716	1.709	0.007
21	1.736	1.737	0.000
22	1.413	1.411	0.002
23	1.736	1.743	0.007
24	1.439	1.433	0.005
25	1.773	1.766	0.007
26	1.757	1.764	0.007
27	1.329	1.334	0.005
28	1.564	1.569	0.005
29	1.318	1.317	0.001
30	1.752	1.743	0.008
31	1.836	1.835	0.001
32	1.507	1.510	0.004
33	1.710	1.709	0.002
34	1.294	1.296	0.003
35	1.397	1.393	0.004
36	1.835	1.839	0.003
37	1.700	1.695	0.005
38	1.813	1.806	0.007
39	1.773	1.766	0.007
40	1.386	1.393	0.007
41	1.825	1.821	0.003
42	1.355	1.360	0.005
43	1.380	1.373	0.007
44	1.820	1.818	0.002
45	1.825	1.825	0.000
46	1.298	1.296	0.002

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
47	1.664	1.662	0.001
48	1.675	1.673	0.002
49	1.523	1.524	0.002
50	1.560	1.556	0.004

Figure 4.4 represents the cell-5 actual and predicted capacity for training and testing split of 70:30% with a total number of instants of 50, using the kNN algorithm.

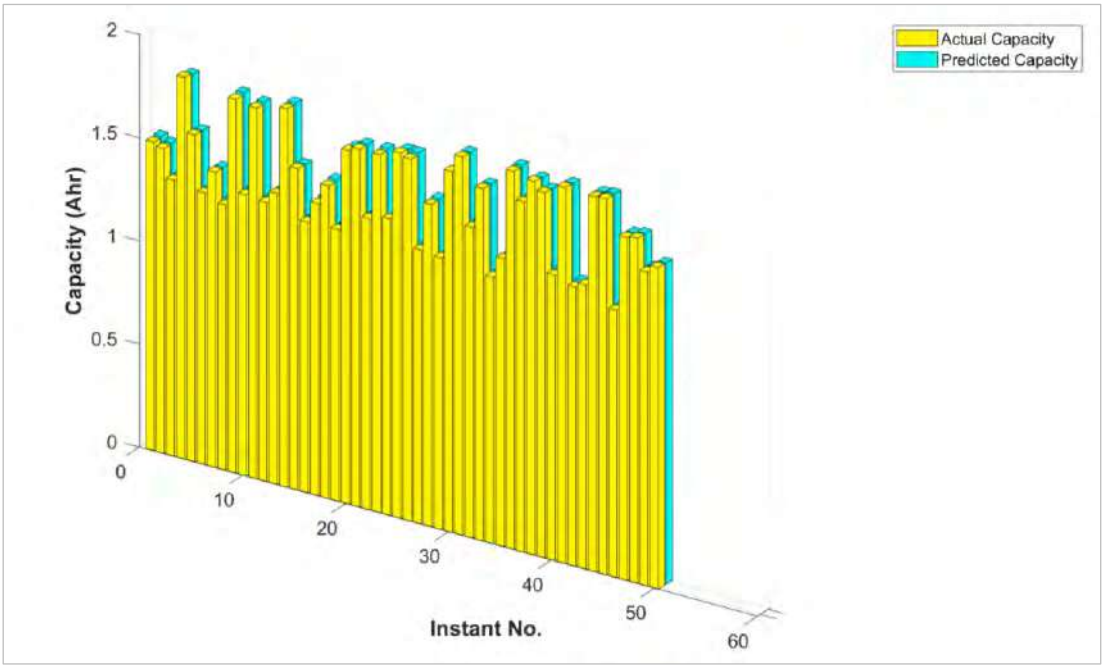


Figure 4.4: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the kNN

Figure 4.5 represents the cell-5 actual and predicted capacity for all studied percentage splits (PSs) using the kNN algorithm.

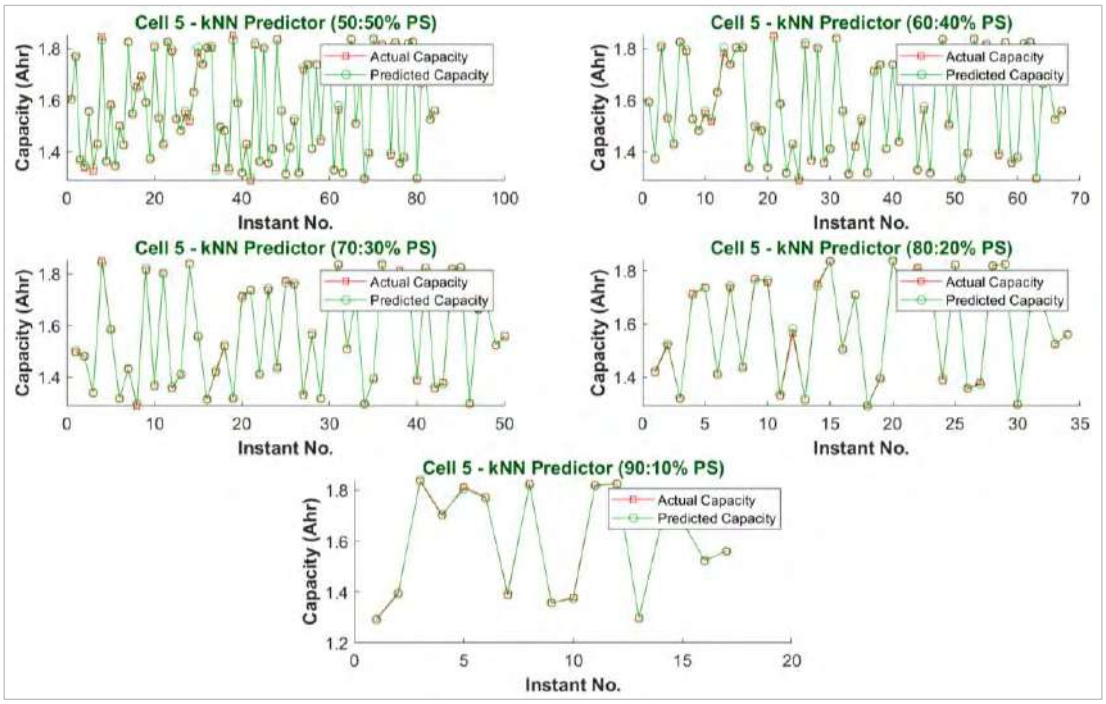


Figure 4.5: Cell-5 actual and predicted capacity for all studied PSs using the kNN

Figure 4.6 shows a zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the kNN algorithm.

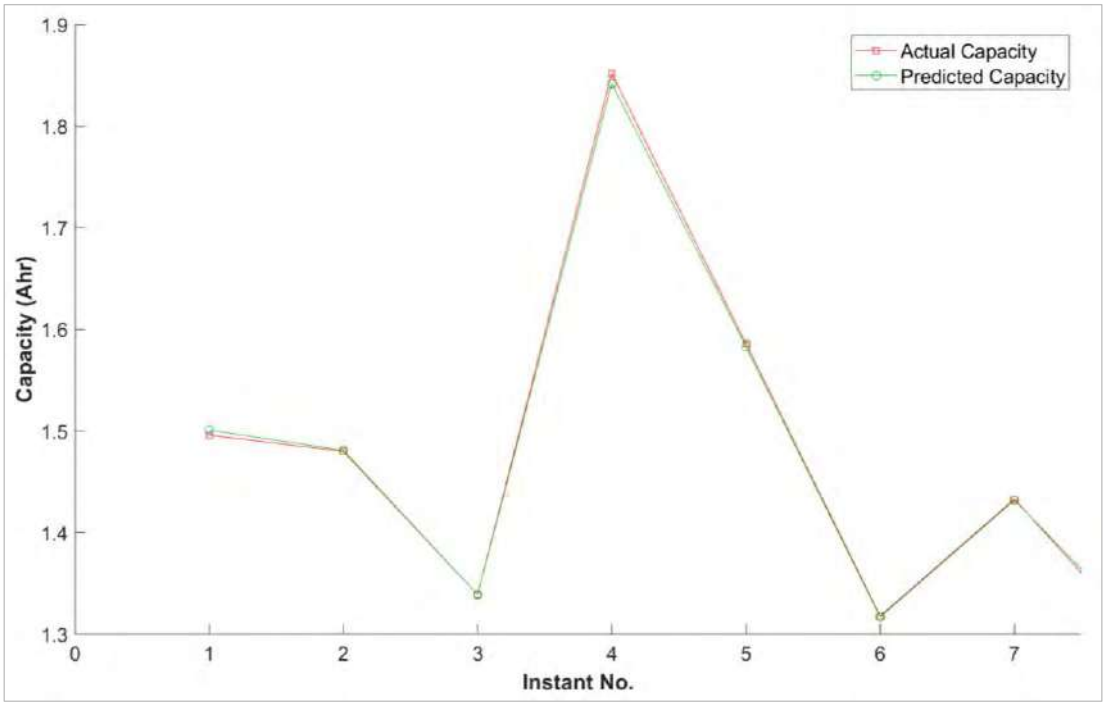


Figure 4.6: Zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the kNN

Figure 4.7 till Figure 4.9 represent the cell-5 actual and predicted capacity for the CV of 3-, 5-, and 10-fold, with a total number of instants in each fold of 56, 34 (or 33), and 17 (or 16), respectively, using the kNN algorithm.

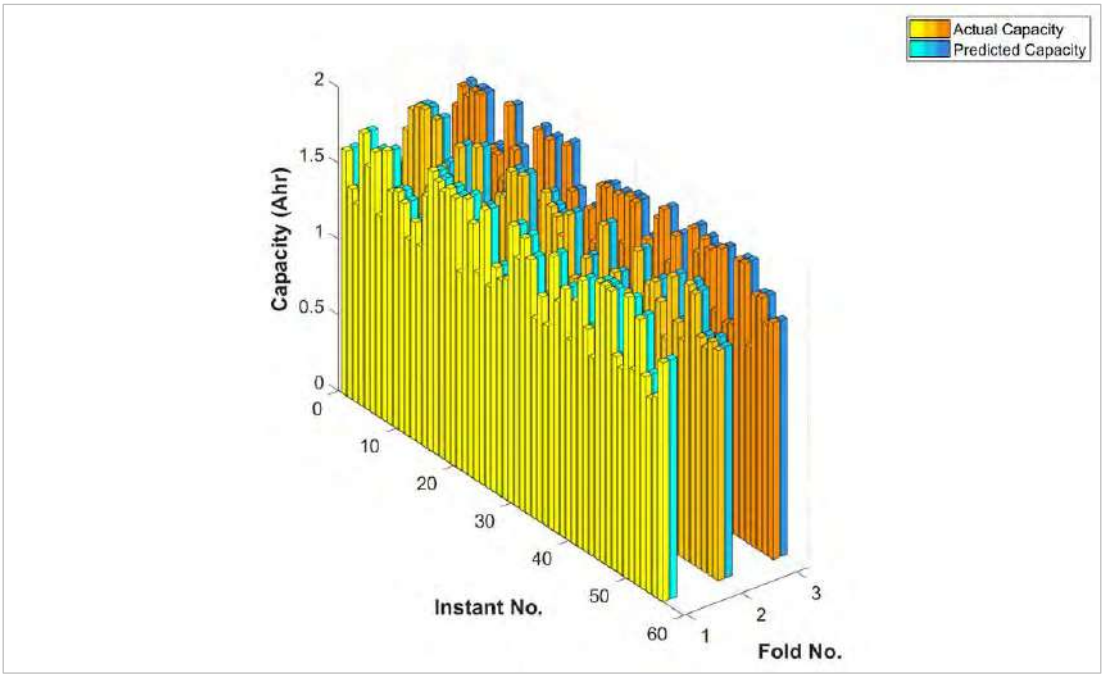


Figure 4.7: Cell-5 actual and predicted capacity for 3-fold CV using the kNN

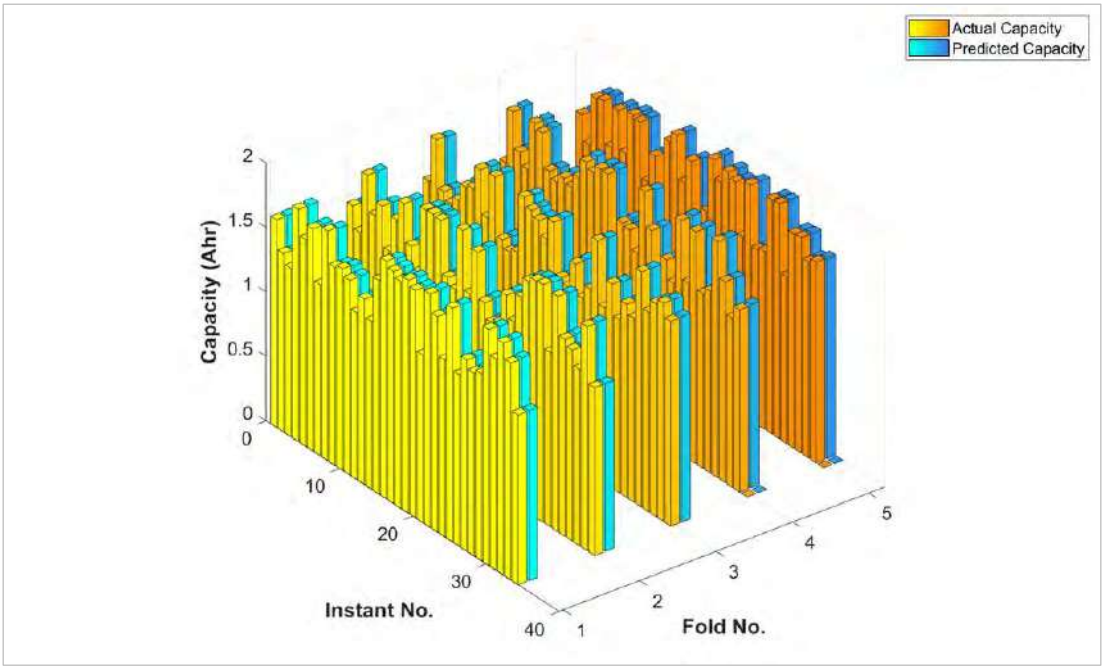


Figure 4.8: Cell-5 actual and predicted capacity for 5-fold CV using the kNN

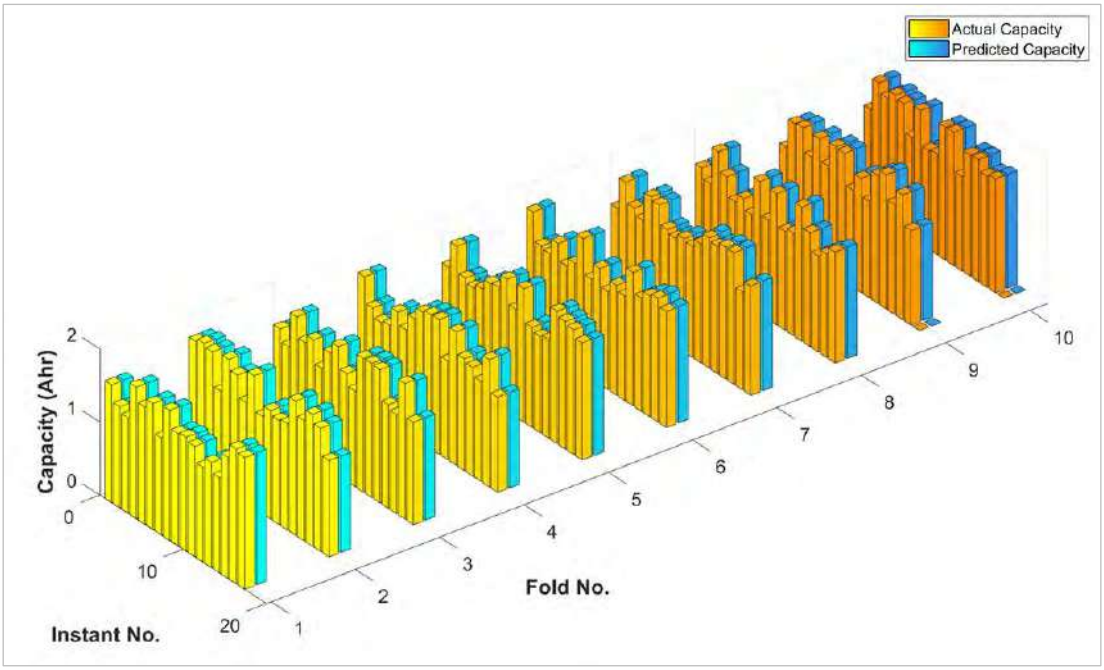


Figure 4.9: Cell-5 actual and predicted capacity for 10-fold CV using the kNN

Figure 4.10 represents the cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the kNN algorithm.

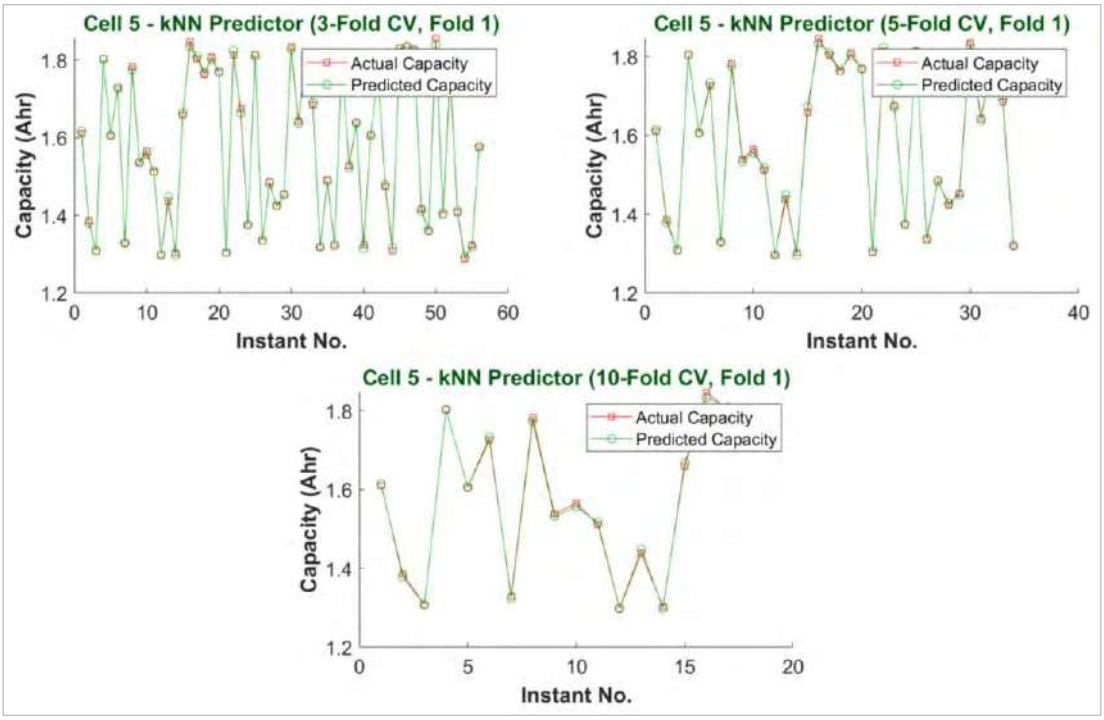


Figure 4.10: Cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the kNN

Table A.1 in Appendix-2 shows the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the kNN algorithm, and the absolute error values, with a total number of instants in this fold of 34.

Figure 4.11 shows a zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the kNN algorithm.

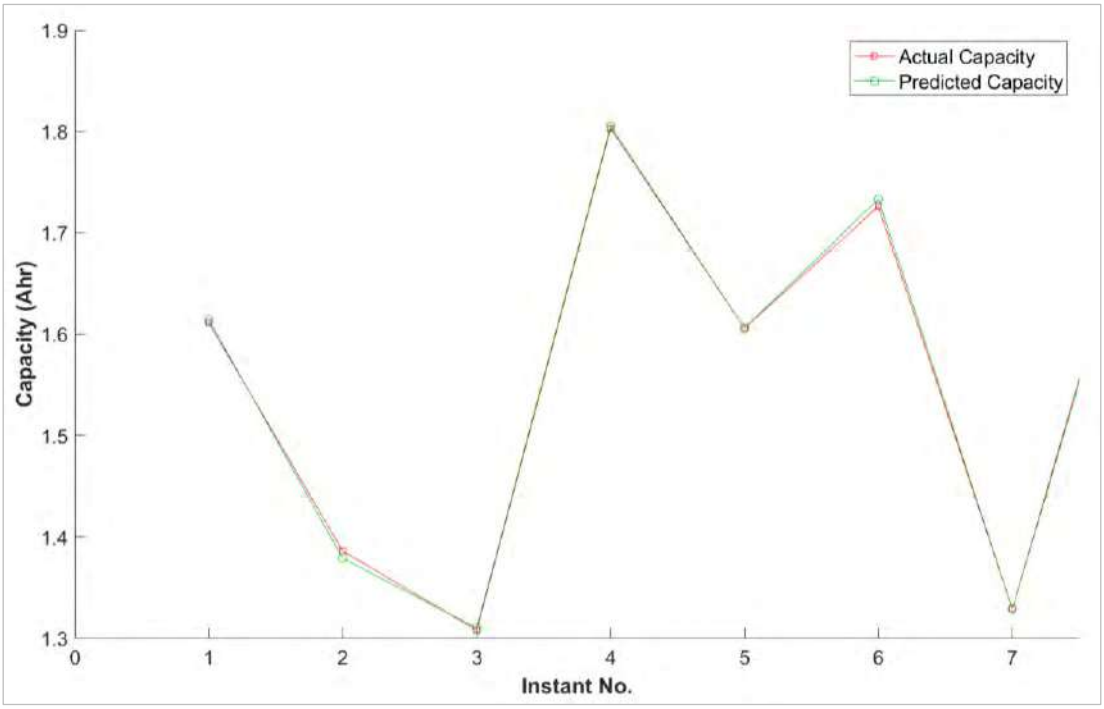


Figure 4.11: Zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the kNN

4.2.2. Artificial Neural Network (ANN)

Table 4.3 shows the cell-5 actual and predicted capacity for training and testing split of 70:30% using the ANN algorithm and the absolute error values, with a total number of instants of 50.

Table 4.3: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the ANN

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.496	1.497	0.001
2	1.480	1.483	0.002
3	1.339	1.334	0.005
4	1.852	1.845	0.007
5	1.586	1.589	0.003

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
6	1.318	1.314	0.004
7	1.433	1.431	0.002
8	1.288	1.285	0.002
9	1.814	1.822	0.007
10	1.365	1.362	0.003
11	1.799	1.802	0.002
12	1.355	1.355	0.000
13	1.412	1.410	0.002
14	1.836	1.840	0.004
15	1.560	1.562	0.002
16	1.313	1.314	0.000
17	1.417	1.421	0.003
18	1.518	1.520	0.002
19	1.319	1.318	0.001
20	1.716	1.722	0.007
21	1.736	1.744	0.008
22	1.413	1.409	0.004
23	1.736	1.739	0.002
24	1.439	1.440	0.001
25	1.773	1.782	0.009
26	1.757	1.759	0.002
27	1.329	1.331	0.002
28	1.564	1.565	0.001
29	1.318	1.315	0.003
30	1.752	1.758	0.006
31	1.836	1.837	0.001
32	1.507	1.507	0.000
33	1.710	1.716	0.005
34	1.294	1.290	0.004
35	1.397	1.396	0.001

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
36	1.835	1.844	0.009
37	1.700	1.705	0.005
38	1.813	1.818	0.004
39	1.773	1.777	0.004
40	1.386	1.389	0.003
41	1.825	1.825	0.001
42	1.355	1.353	0.002
43	1.380	1.379	0.002
44	1.820	1.819	0.001
45	1.825	1.830	0.005
46	1.298	1.292	0.006
47	1.664	1.670	0.006
48	1.675	1.682	0.007
49	1.523	1.525	0.002
50	1.560	1.560	0.000

Figure 4.12 represents the cell-5 actual and predicted capacity for training and testing split of 70:30% with a total number of instants of 50, using the ANN algorithm.

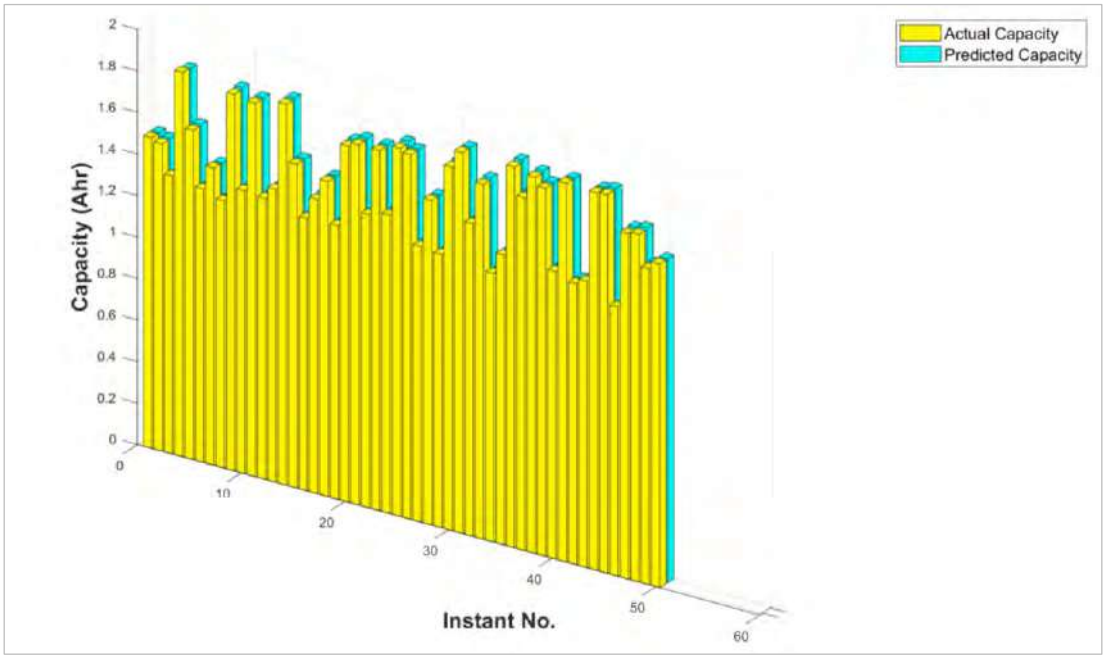


Figure 4.12: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the ANN

Figure 4.13 represents the cell-5 actual and predicted capacity for all studied PSs using the ANN algorithm.

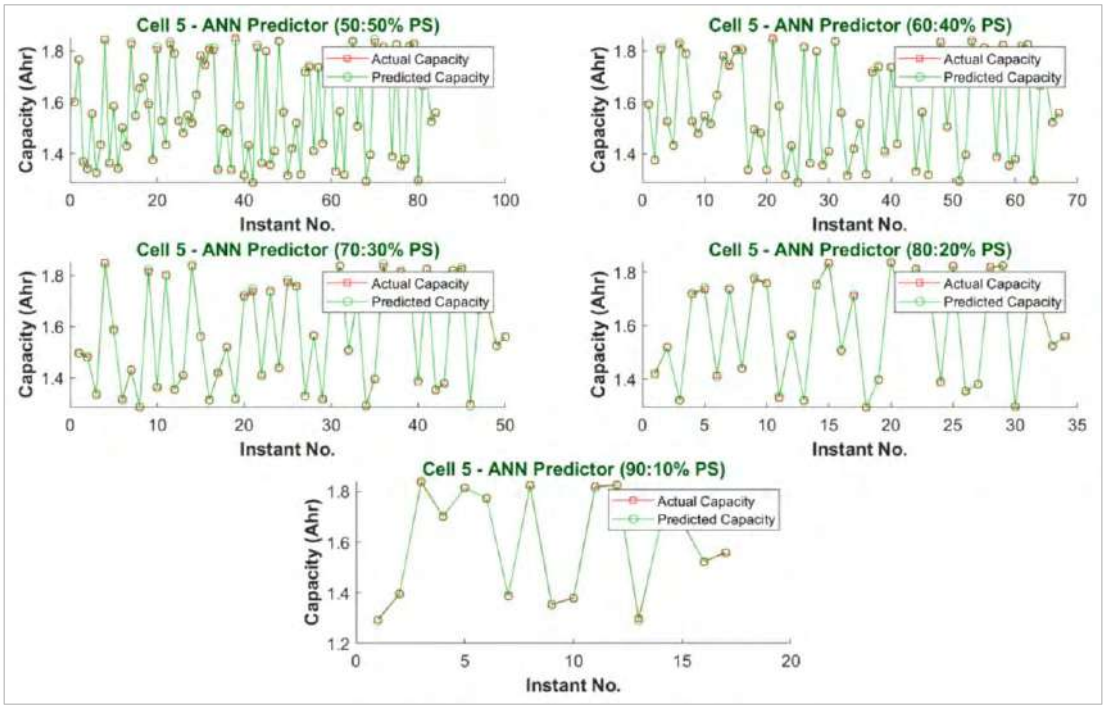


Figure 4.13: Cell-5 actual and predicted capacity for all studied PSs using the ANN

Figure 4.14 shows a zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the ANN algorithm.

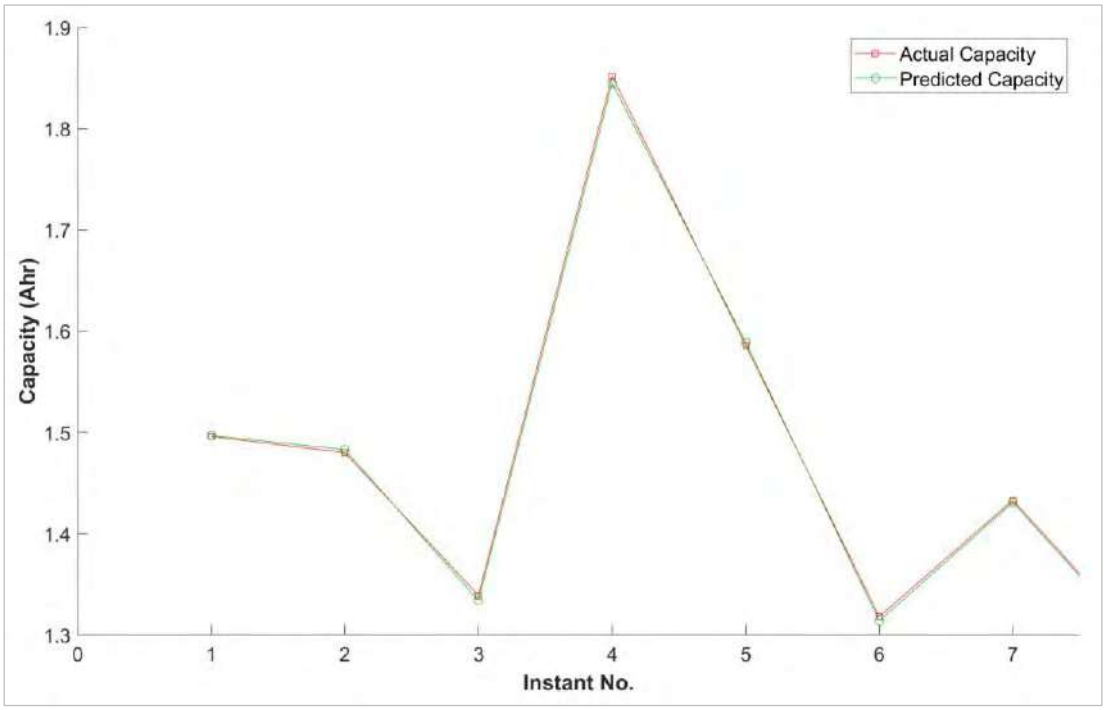


Figure 4.14: Zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the ANN

Figure 4.15 till Figure 4.17 represent the cell-5 actual and predicted capacity for The CV of 3-, 5-, and 10-fold, with a total number of instants in each fold of 56, 34 (or 33), and 17 (or 16), respectively, using the ANN algorithm.

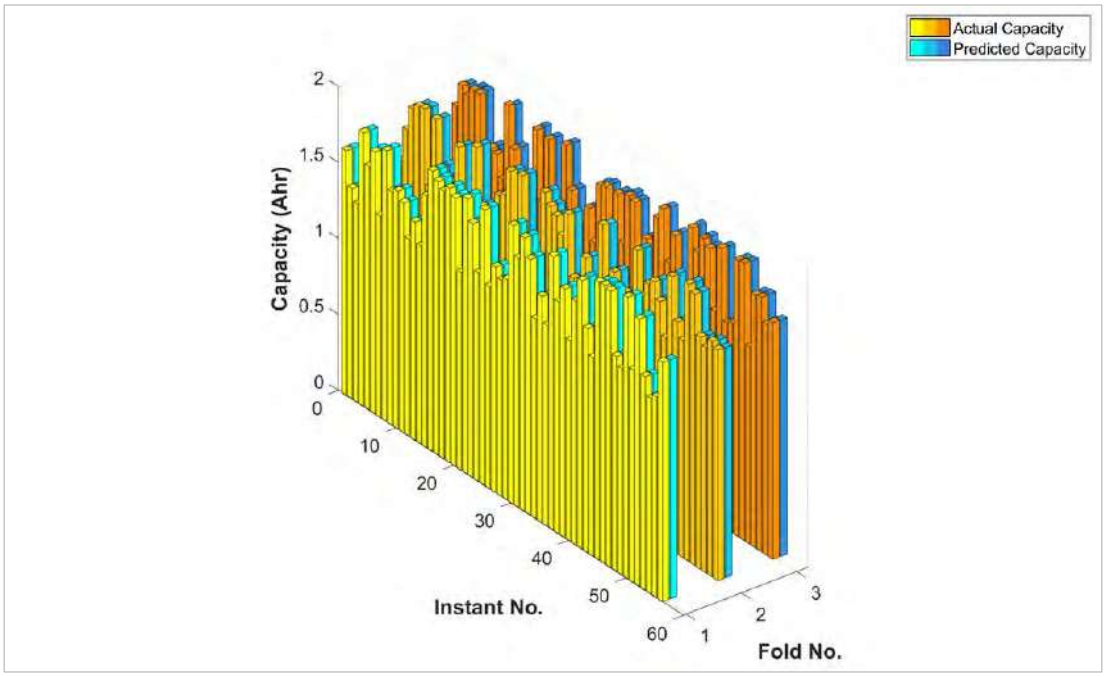


Figure 4.15: Cell-5 actual and predicted capacity for 3-fold CV using the ANN

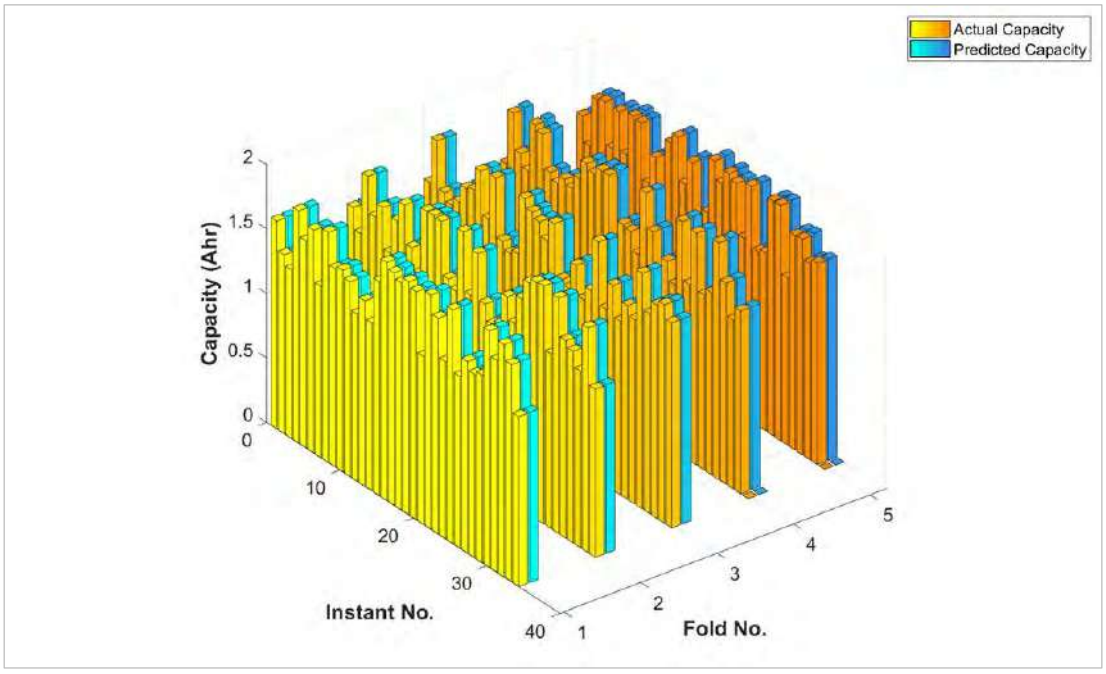


Figure 4.16: Cell-5 actual and predicted capacity for 5-fold CV using the ANN

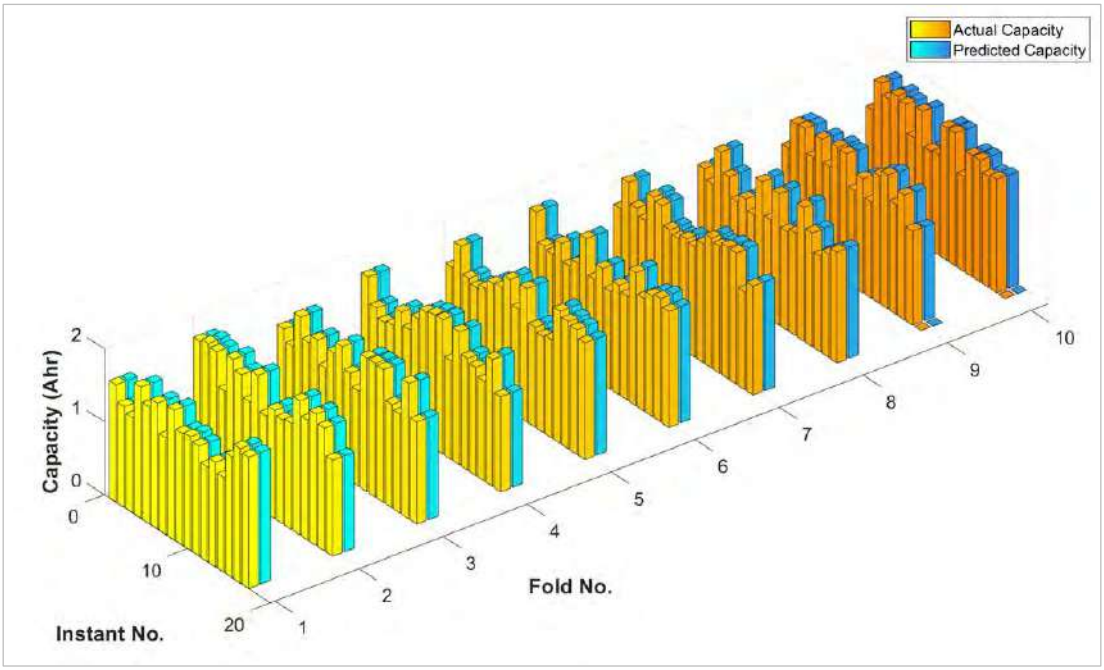


Figure 4.17: Cell-5 actual and predicted capacity for 10-fold CV using the ANN

Figure 4.18 represents the cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the ANN algorithm.

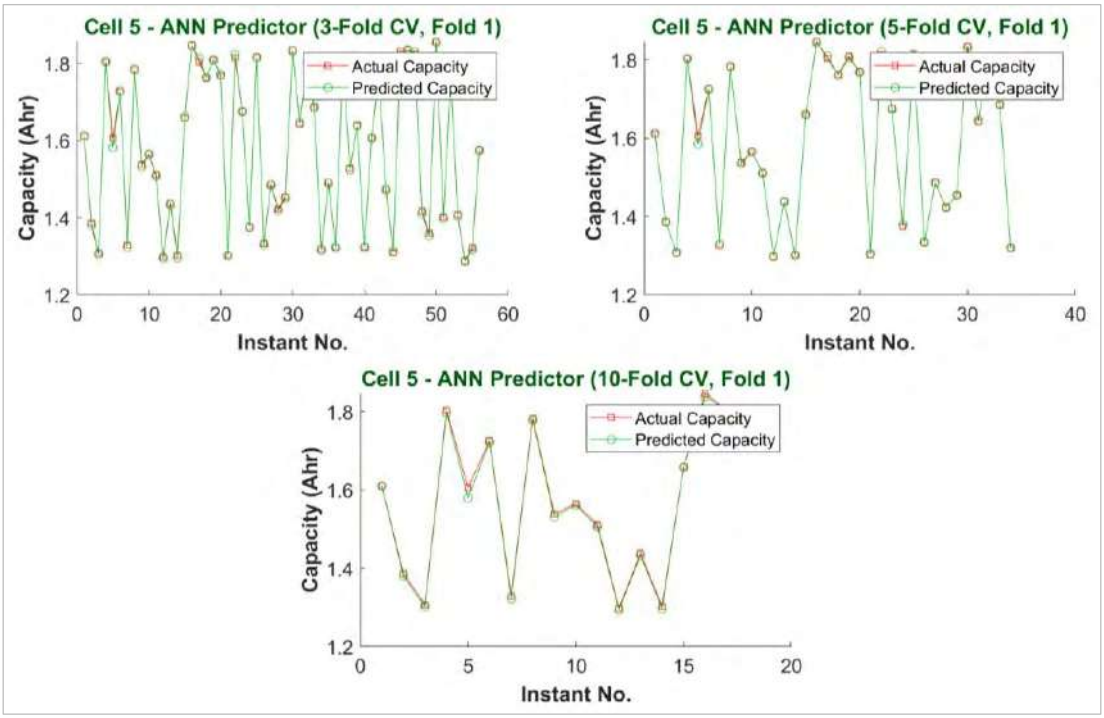


Figure 4.18: Cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the ANN

Table A.2 in Appendix-2 shows the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the ANN algorithm, and the absolute error values, with a total number of instants in this fold of 34.

Figure 4.19 shows a zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the ANN algorithm.

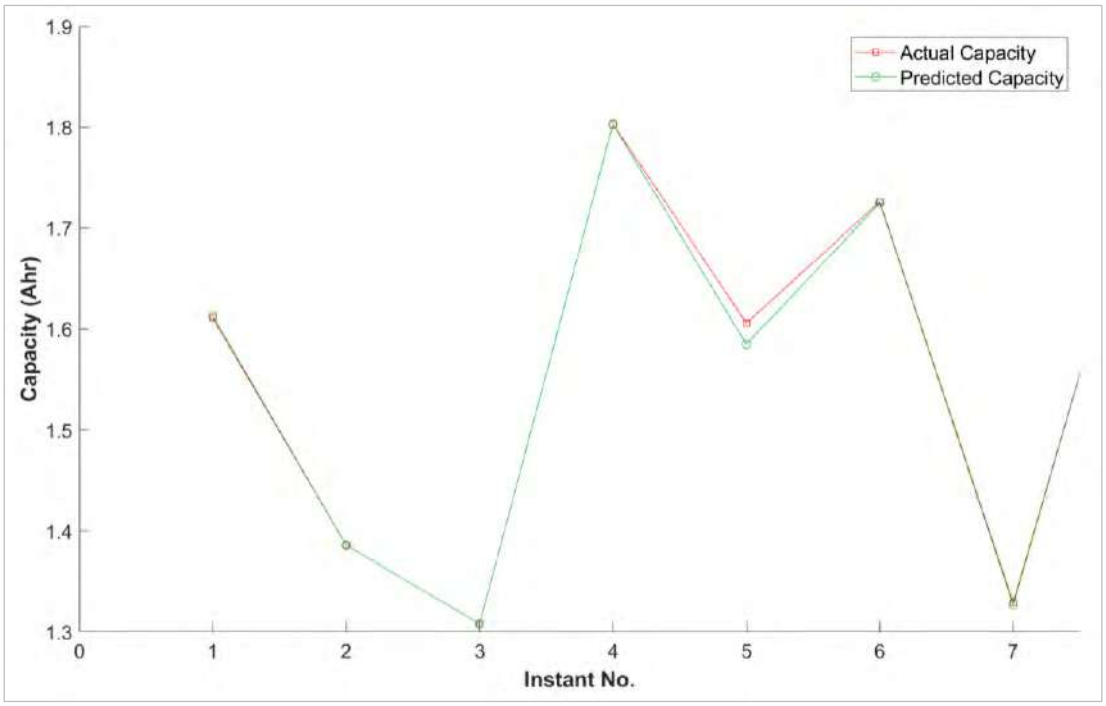


Figure 4.19: Zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the ANN

4.2.3. Linear Regression (LR)

Table 4.4 shows the cell-5 actual and predicted capacity for training and testing split of 70:30% using the LR algorithm and the absolute error values, with a total number of instants of 50.

Table 4.4: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the LR

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.496	1.497	0.000
2	1.480	1.482	0.001
3	1.339	1.336	0.003
4	1.852	1.847	0.005
5	1.586	1.585	0.001

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
6	1.318	1.316	0.002
7	1.433	1.431	0.003
8	1.288	1.288	0.001
9	1.814	1.817	0.003
10	1.365	1.363	0.002
11	1.799	1.799	0.000
12	1.355	1.356	0.001
13	1.412	1.411	0.002
14	1.836	1.836	0.001
15	1.560	1.559	0.001
16	1.313	1.316	0.003
17	1.417	1.420	0.003
18	1.518	1.518	0.000
19	1.319	1.320	0.002
20	1.716	1.718	0.002
21	1.736	1.738	0.002
22	1.413	1.409	0.004
23	1.736	1.735	0.001
24	1.439	1.440	0.002
25	1.773	1.776	0.003
26	1.757	1.757	0.000
27	1.329	1.333	0.004
28	1.564	1.563	0.001
29	1.318	1.318	0.000
30	1.752	1.753	0.001
31	1.836	1.832	0.004
32	1.507	1.505	0.001
33	1.710	1.711	0.001
34	1.294	1.293	0.000
35	1.397	1.397	0.000

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
36	1.835	1.839	0.004
37	1.700	1.701	0.001
38	1.813	1.812	0.001
39	1.773	1.771	0.002
40	1.386	1.390	0.004
41	1.825	1.820	0.005
42	1.355	1.353	0.002
43	1.380	1.380	0.001
44	1.820	1.817	0.002
45	1.825	1.825	0.000
46	1.298	1.296	0.003
47	1.664	1.666	0.003
48	1.675	1.677	0.003
49	1.523	1.523	0.001
50	1.560	1.557	0.003

Figure 4.20 represents the cell-5 actual and predicted capacity for training and testing split of 70:30% with a total number of instants of 50, using the LR algorithm.

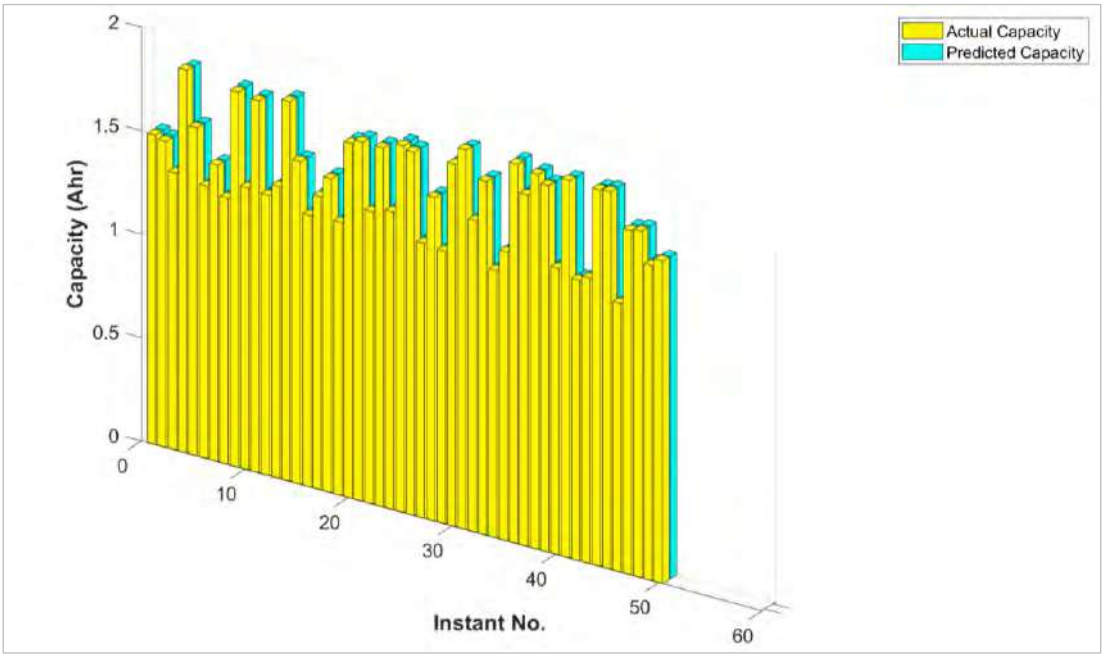


Figure 4.20: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the LR

Figure 4.21 represents the cell-5 actual and predicted capacity for all studied PSs using the LR algorithm.

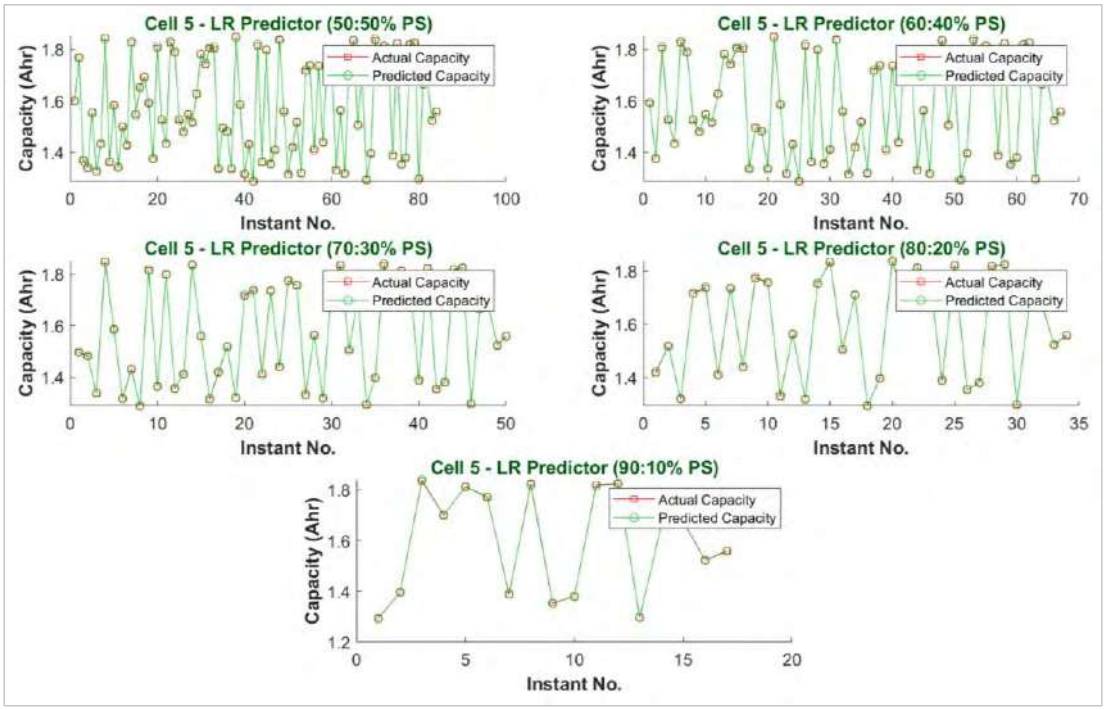


Figure 4.21: Cell-5 actual and predicted capacity for all studied PSs using the LR

Figure 4.22 shows a zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the LR algorithm.

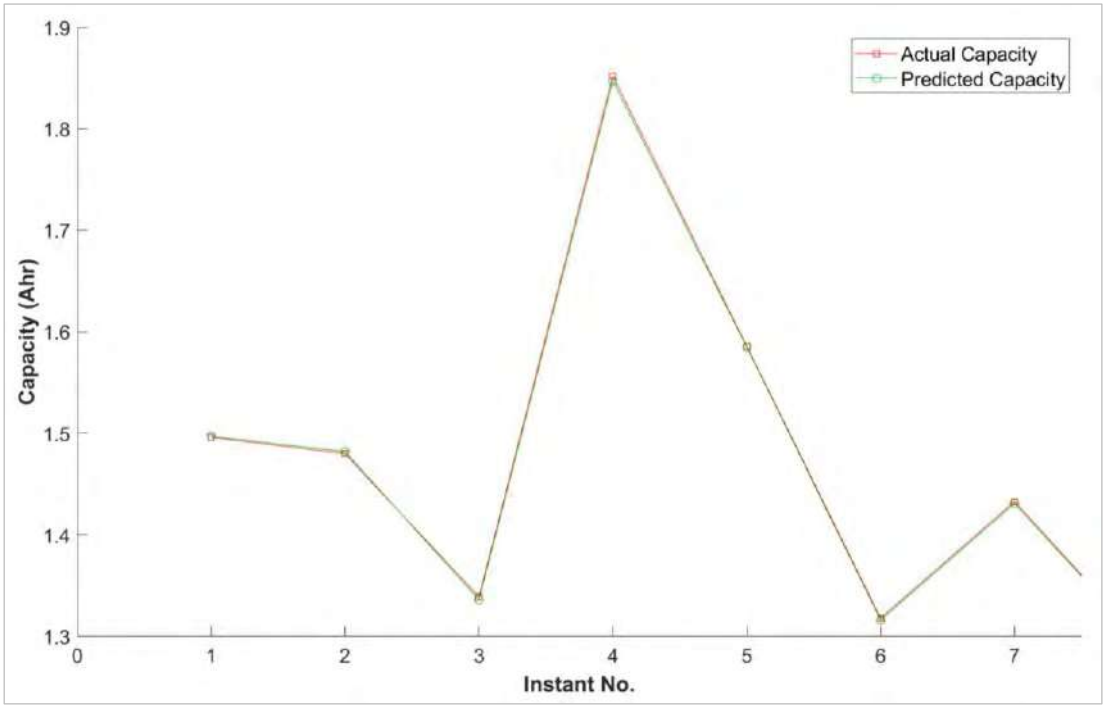


Figure 4.22: Zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the LR

Figure 4.23 till Figure 4.25 represent the cell-5 actual and predicted capacity for The CV of 3-, 5-, and 10-fold, with a total number of instants in each fold of 56, 34 (or 33), and 17 (or 16), respectively, using the LR algorithm.

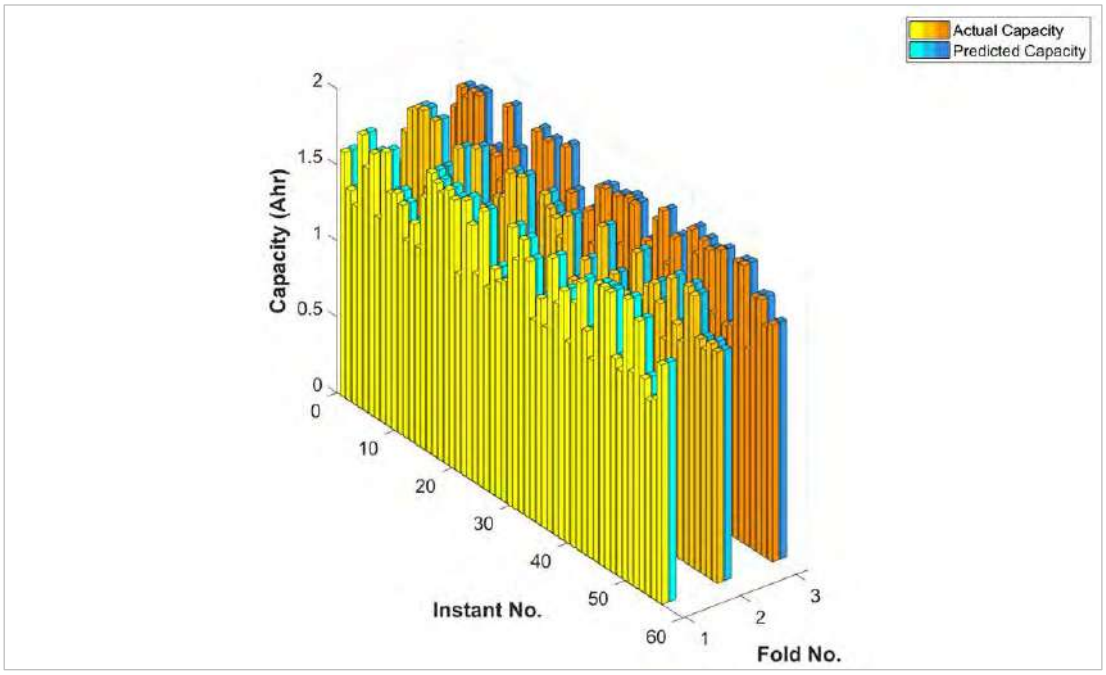


Figure 4.23: Cell-5 actual and predicted capacity for 3-fold CV using the LR

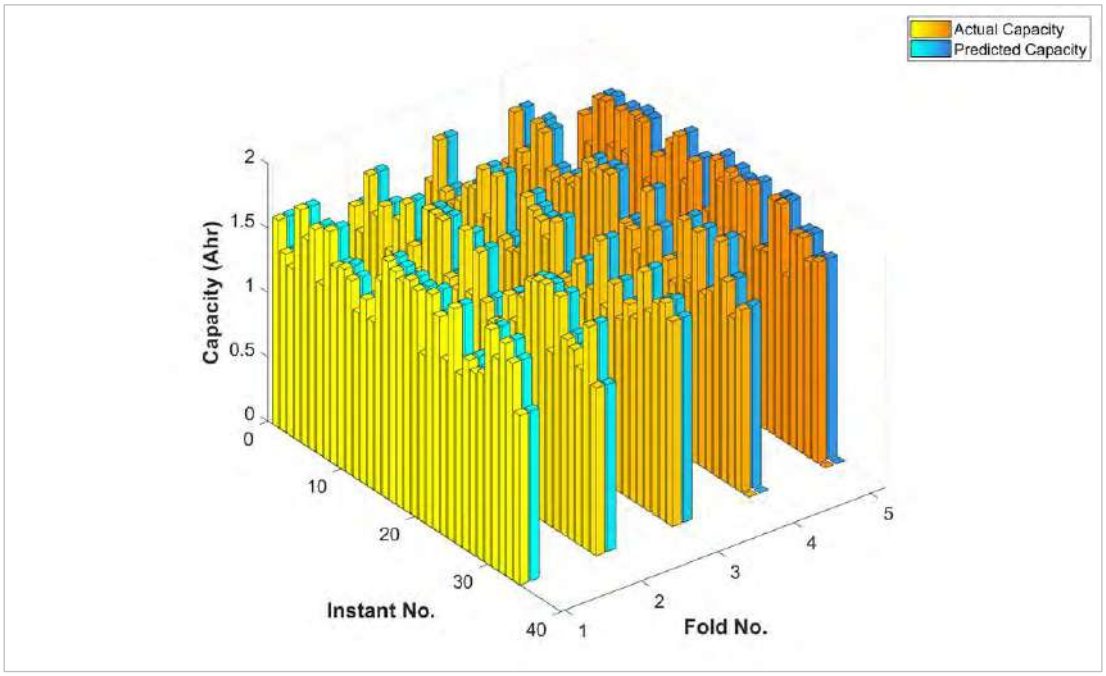


Figure 4.24: Cell-5 actual and predicted capacity for 5-fold CV using the LR

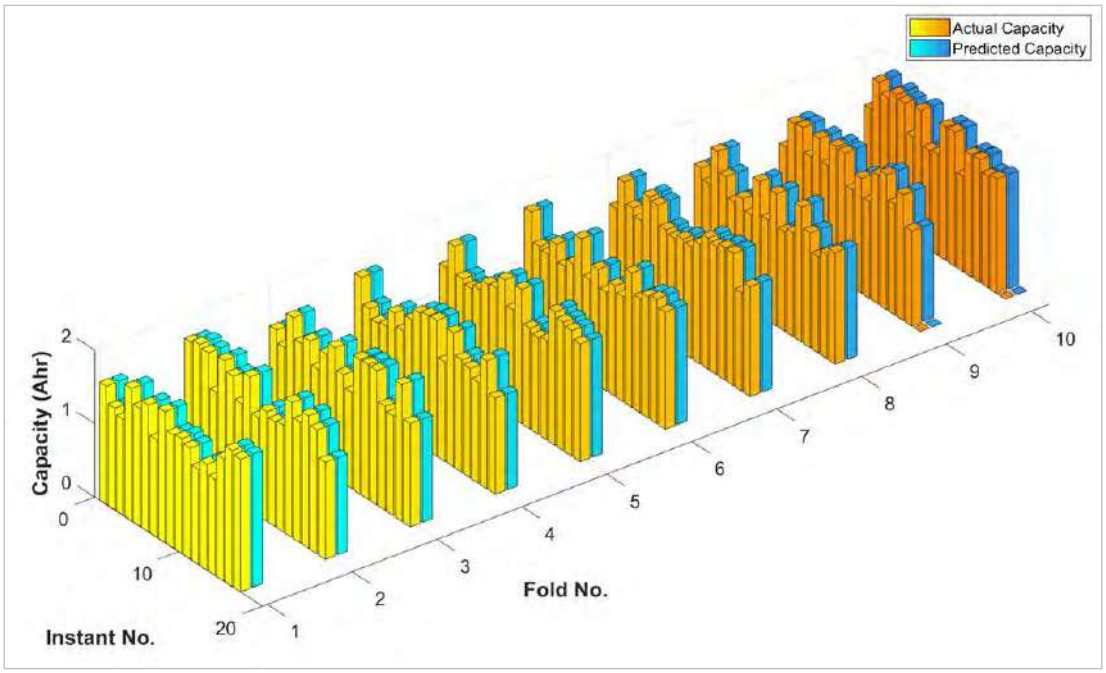


Figure 4.25: Cell-5 actual and predicted capacity for 10-fold CV using the LR

Figure 4.26 represents the cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the LR algorithm.

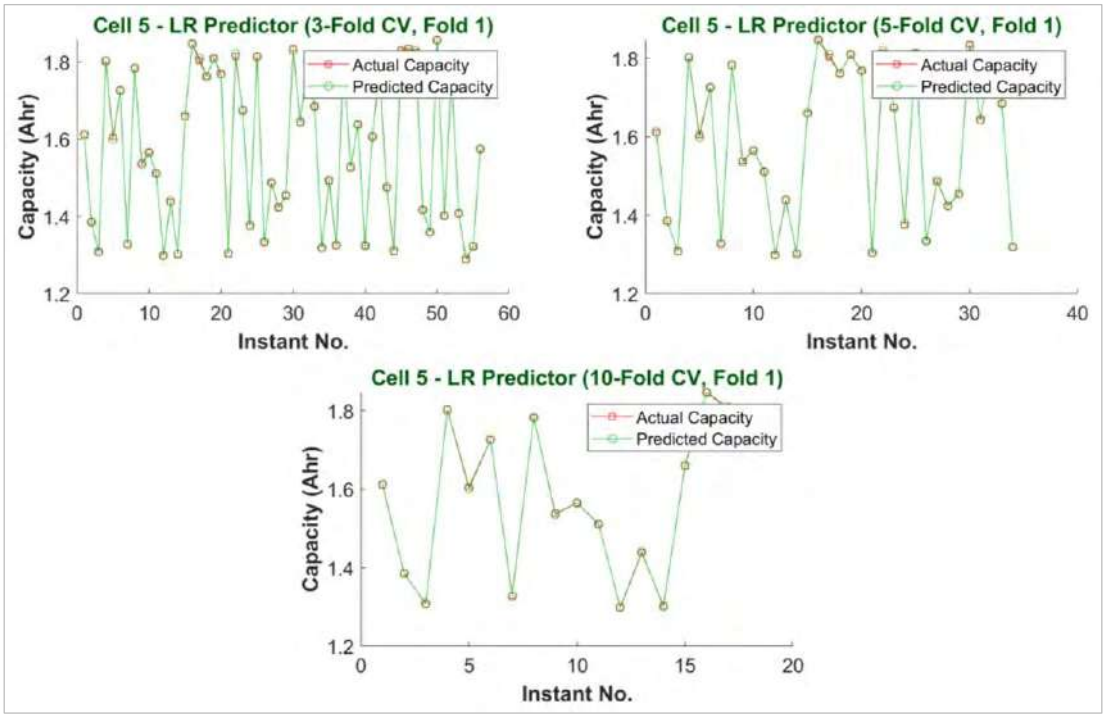


Figure 4.26: Cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the LR

Table A.3 in Appendix-2 shows the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the LR algorithm, and the absolute error values, with a total number of instants in this fold of 34.

Figure 4.27 shows a zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the LR algorithm.

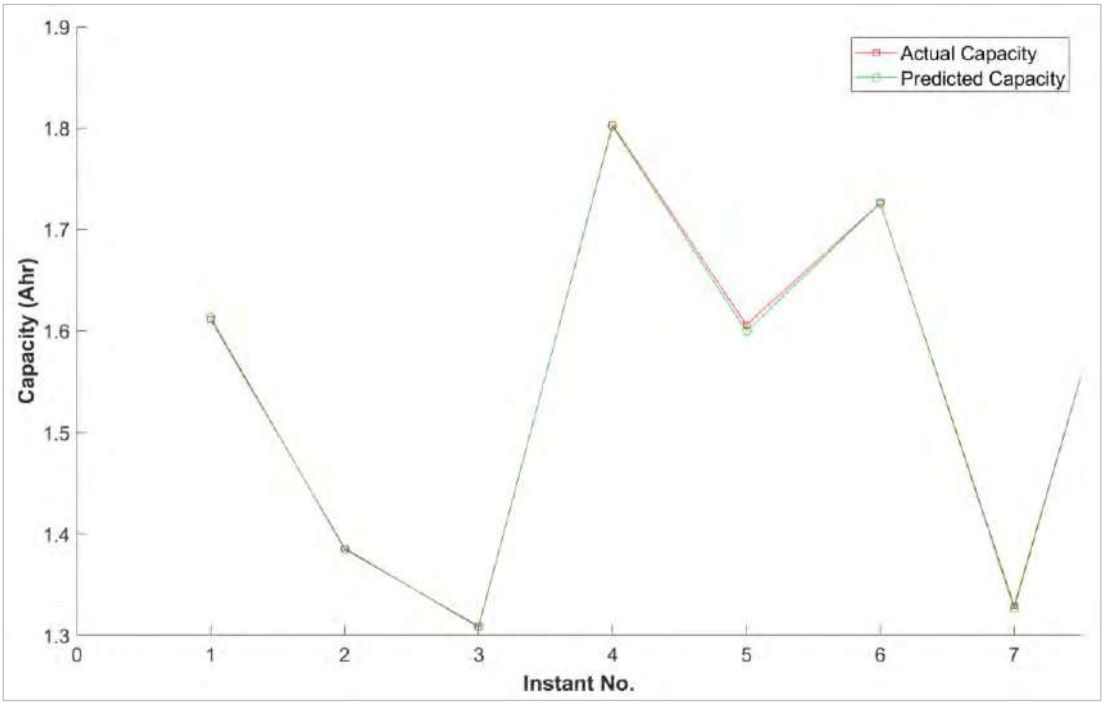


Figure 4.27: Zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the LR

4.2.4. Random Tree (RT)

Table 4.5 shows the cell-5 actual and predicted capacity for training and testing split of 70:30% using the RT algorithm and the absolute error values, with a total number of instants of 50.

Table 4.5: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the RT

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.496	1.514	0.018
2	1.480	1.483	0.003
3	1.339	1.343	0.004
4	1.852	1.849	0.002
5	1.586	1.590	0.004
6	1.318	1.307	0.012

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
7	1.433	1.431	0.002
8	1.288	1.293	0.005
9	1.814	1.814	0.000
10	1.365	1.360	0.005
11	1.799	1.803	0.003
12	1.355	1.360	0.005
13	1.412	1.408	0.004
14	1.836	1.849	0.013
15	1.560	1.551	0.009
16	1.313	1.307	0.007
17	1.417	1.431	0.014
18	1.518	1.514	0.003
19	1.319	1.324	0.005
20	1.716	1.726	0.010
21	1.736	1.743	0.007
22	1.413	1.408	0.004
23	1.736	1.743	0.007
24	1.439	1.431	0.007
25	1.773	1.766	0.007
26	1.757	1.766	0.009
27	1.329	1.330	0.001
28	1.564	1.570	0.006
29	1.318	1.324	0.005
30	1.752	1.743	0.008
31	1.836	1.830	0.006
32	1.507	1.514	0.008
33	1.710	1.706	0.005
34	1.294	1.293	0.001
35	1.397	1.389	0.008
36	1.835	1.830	0.006

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
37	1.700	1.706	0.006
38	1.813	1.814	0.001
39	1.773	1.766	0.007
40	1.386	1.408	0.022
41	1.825	1.814	0.010
42	1.355	1.360	0.005
43	1.380	1.372	0.008
44	1.820	1.814	0.006
45	1.825	1.830	0.005
46	1.298	1.293	0.005
47	1.664	1.656	0.007
48	1.675	1.675	0.000
49	1.523	1.531	0.008
50	1.560	1.551	0.009

Figure 4.28 represents the cell-5 actual and predicted capacity for training and testing split of 70:30% with a total number of instants of 50, using the RT algorithm.

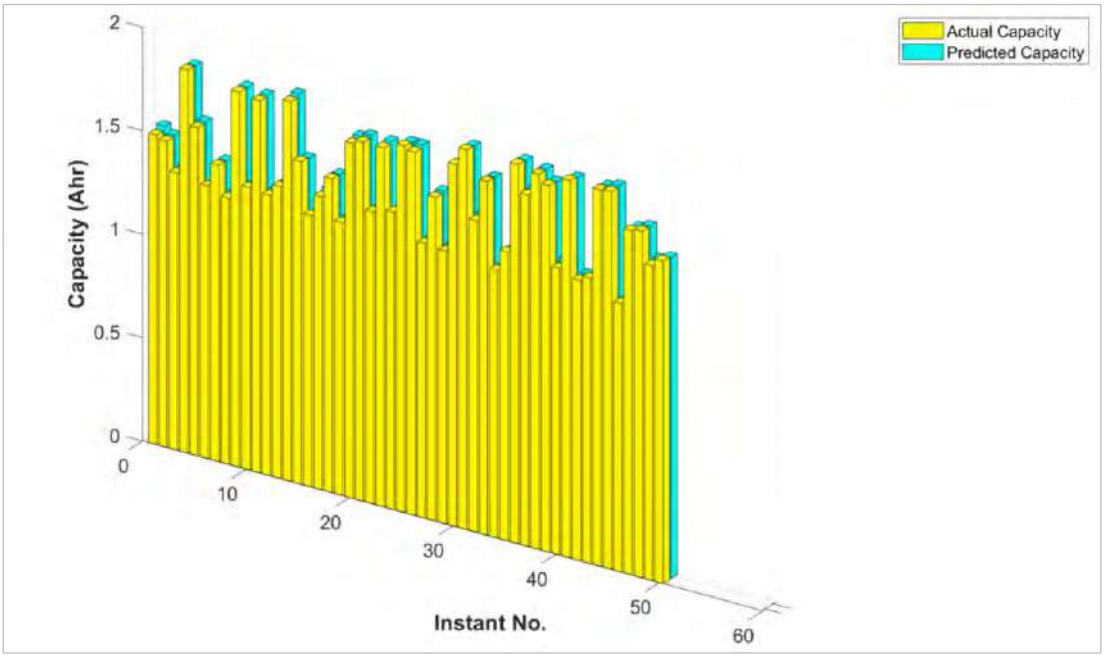


Figure 4.28: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the RT

Figure 4.29 represents the cell-5 actual and predicted capacity for all studied PSs using the RT algorithm.

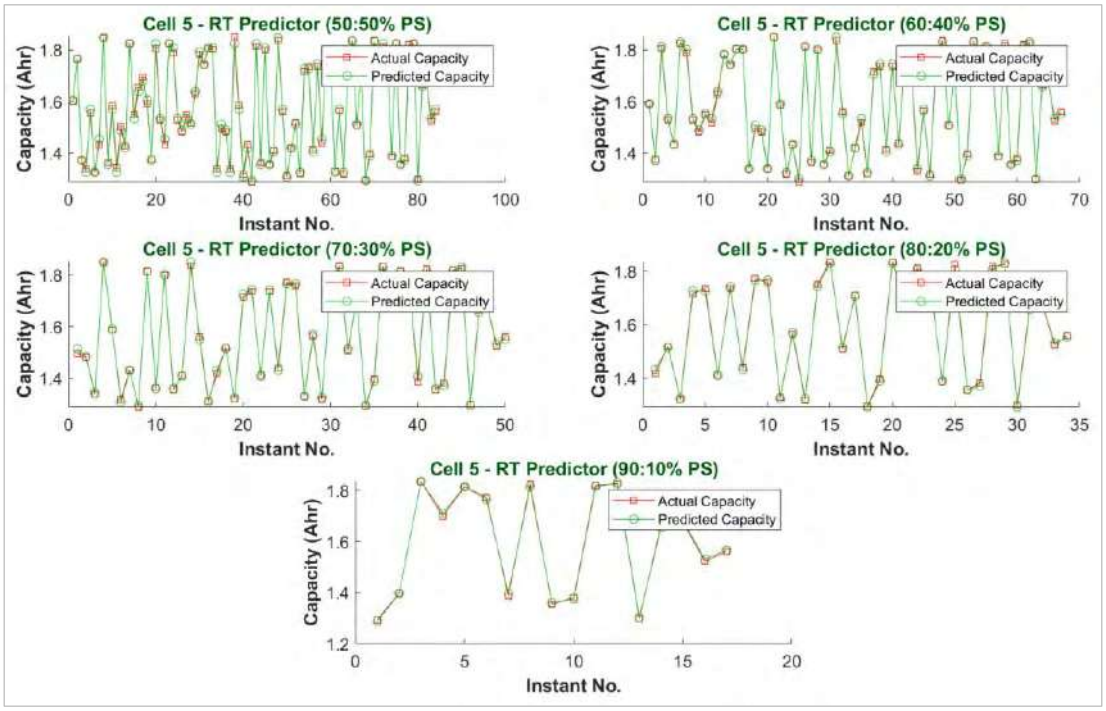


Figure 4.29: Cell-5 actual and predicted capacity for all studied PSs using the RT

Figure 4.30 shows a zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the RT algorithm.

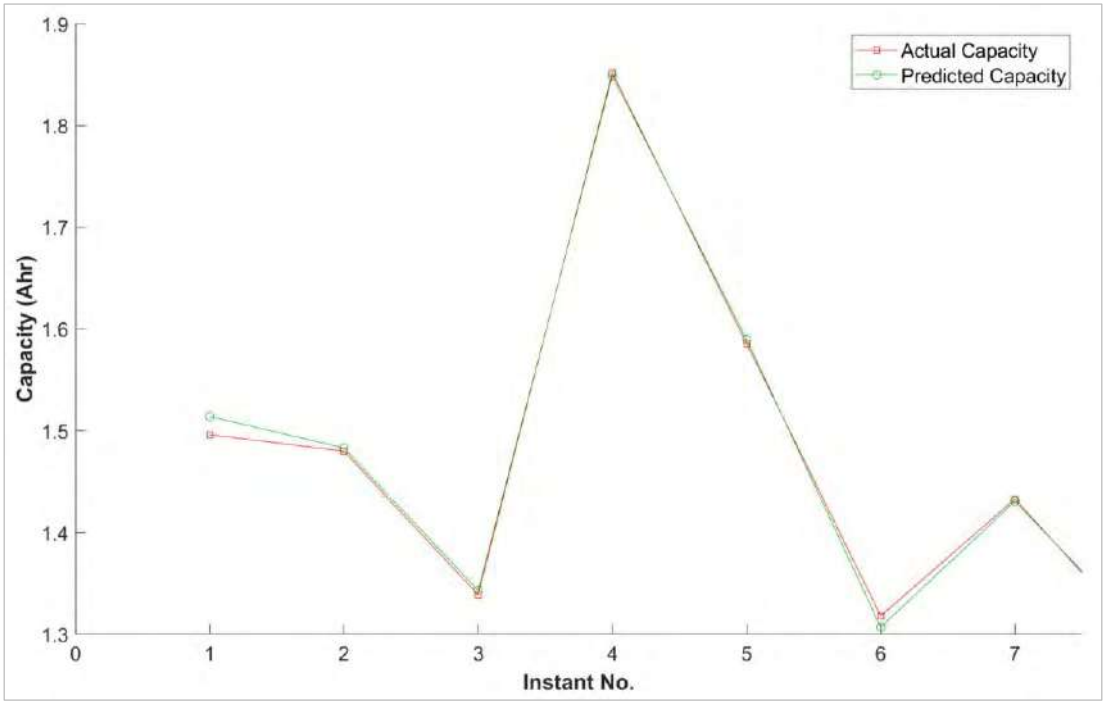


Figure 4.30: Zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the RT

Figure 4.31 till Figure 4.33 represent the cell-5 actual and predicted capacity for The CV of 3-, 5-, and 10-fold, with a total number of instants in each fold of 56, 34 (or 33), and 17 (or 16), respectively, using the RT algorithm.

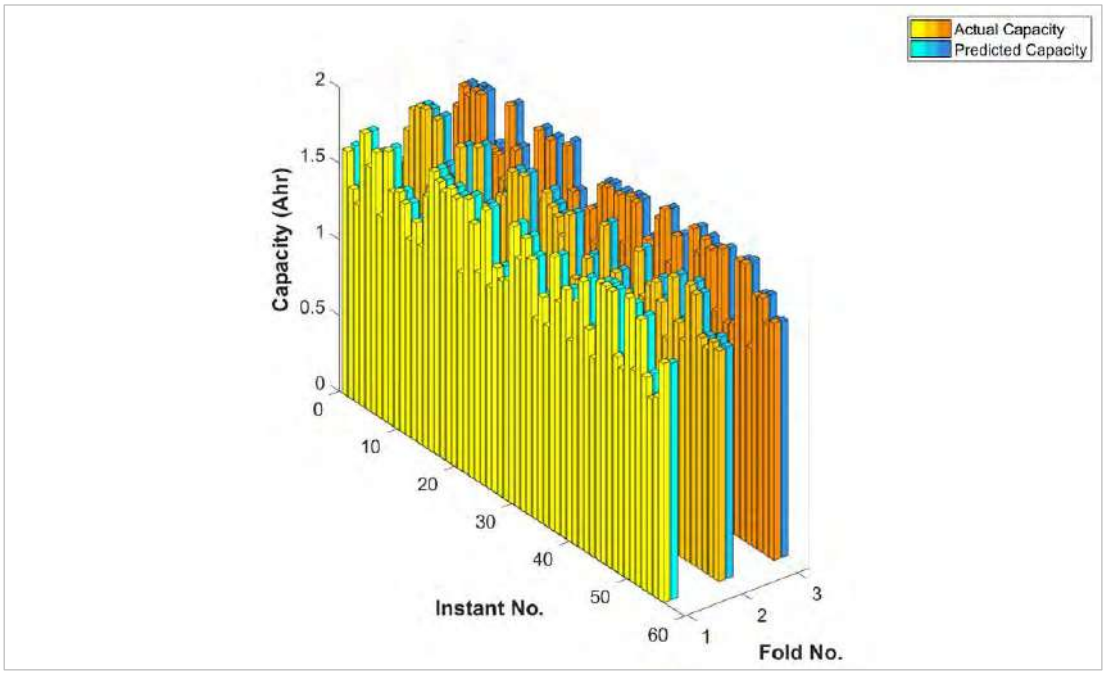


Figure 4.31: Cell-5 actual and predicted capacity for 3-fold CV using the RT

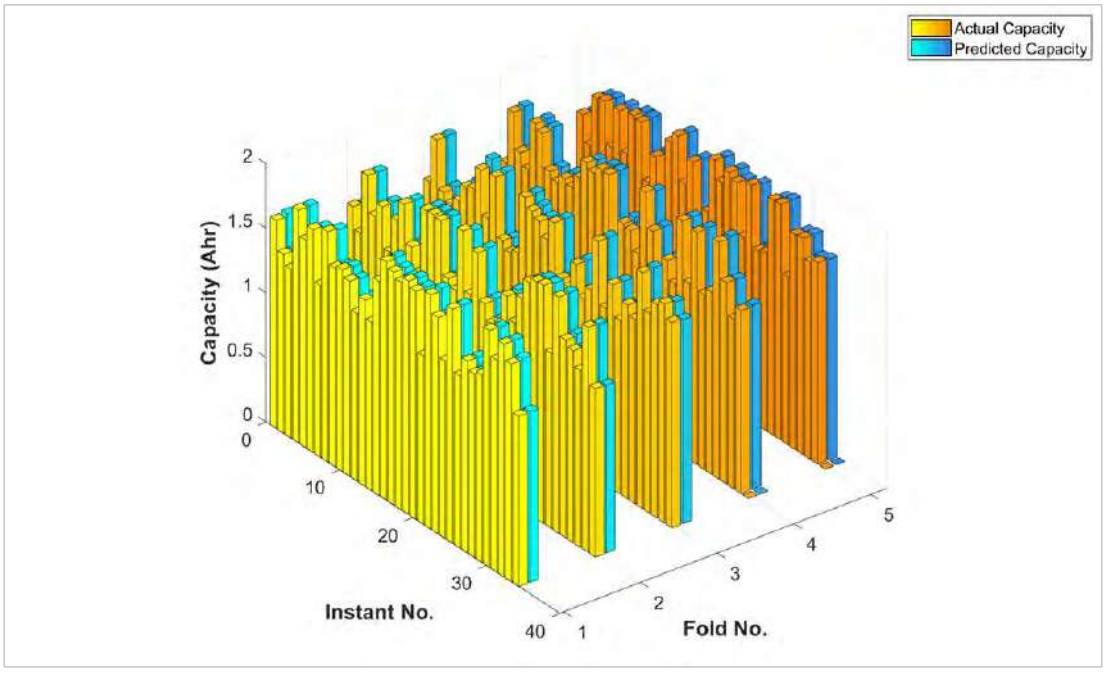


Figure 4.32: Cell-5 actual and predicted capacity for 5-fold CV using the RT

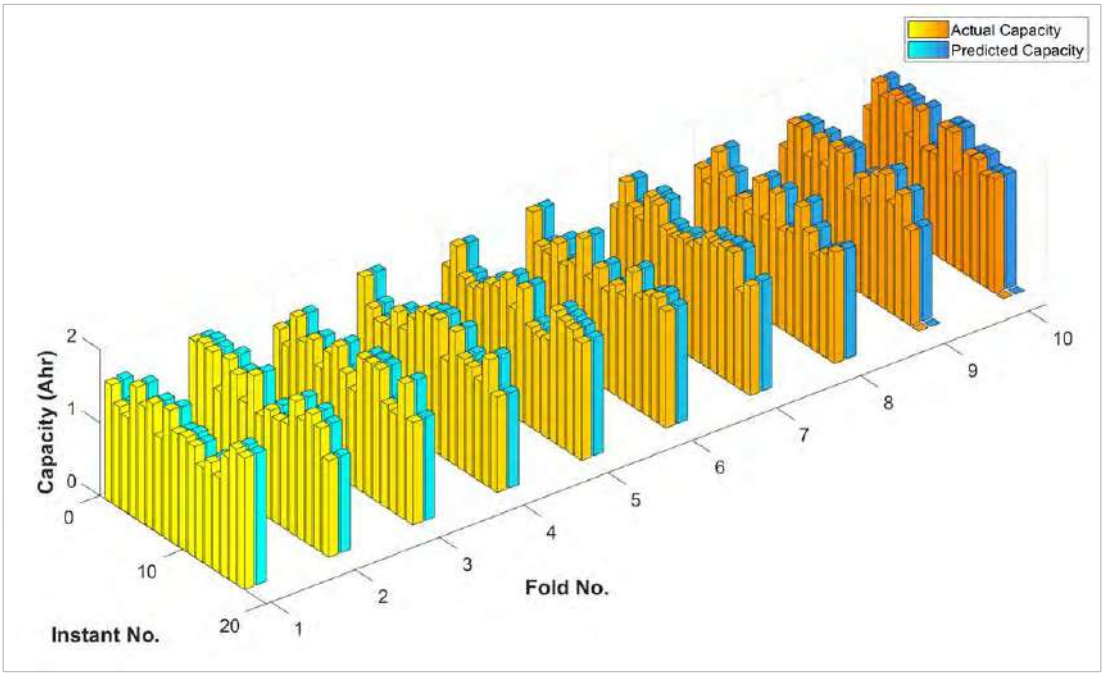


Figure 4.33: Cell-5 actual and predicted capacity for 10-fold CV using the RT

Figure 4.34 represents the cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the RT algorithm.

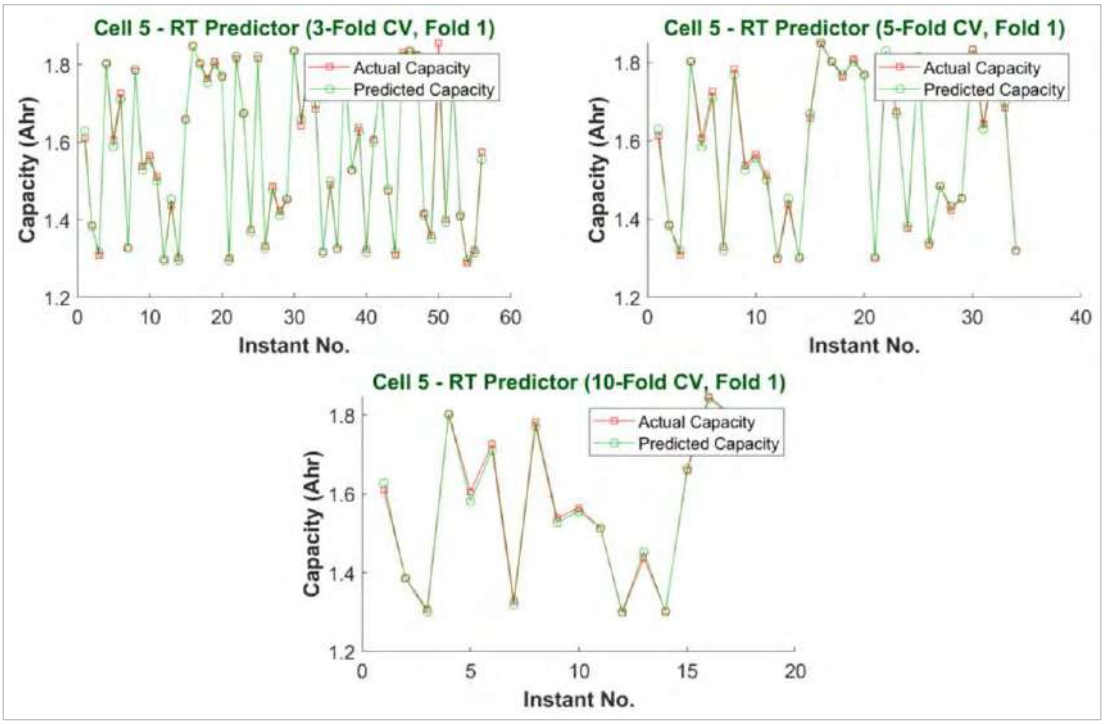


Figure 4.34: Cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the RT

Table A.4 in Appendix-2 shows the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RT algorithm, and the absolute error values, with a total number of instants in this fold of 34.

Figure 4.35 shows a zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RT algorithm.

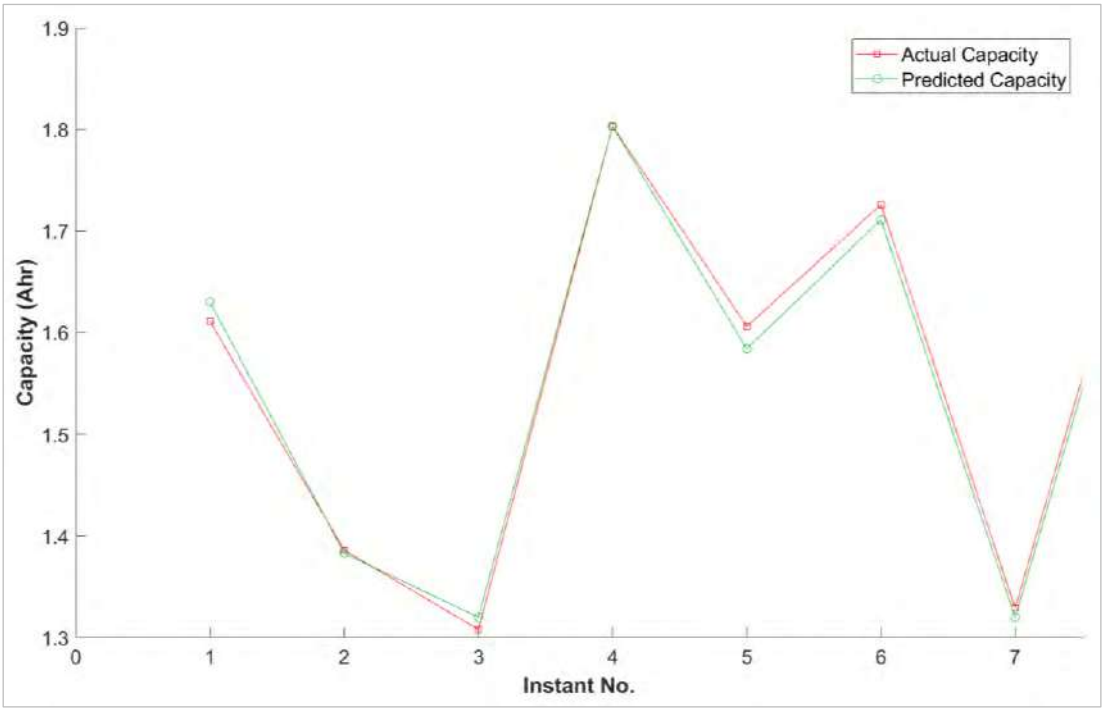


Figure 4.35: Zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RT

4.2.5. Random Forest (RF)

Table 4.6 shows the cell-5 actual and predicted capacity for training and testing split of 70:30% using the RF algorithm and the absolute error values, with a total number of instants of 50.

Table 4.6: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the RF

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.496	1.503	0.006
2	1.480	1.483	0.003
3	1.339	1.336	0.003
4	1.852	1.843	0.009
5	1.586	1.590	0.004
6	1.318	1.315	0.004

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
7	1.433	1.431	0.002
8	1.288	1.296	0.008
9	1.814	1.812	0.002
10	1.365	1.368	0.003
11	1.799	1.805	0.006
12	1.355	1.359	0.004
13	1.412	1.408	0.004
14	1.836	1.845	0.009
15	1.560	1.553	0.007
16	1.313	1.312	0.002
17	1.417	1.423	0.006
18	1.518	1.521	0.004
19	1.319	1.320	0.001
20	1.716	1.722	0.006
21	1.736	1.743	0.007
22	1.413	1.407	0.005
23	1.736	1.742	0.006
24	1.439	1.434	0.005
25	1.773	1.765	0.008
26	1.757	1.764	0.007
27	1.329	1.332	0.003
28	1.564	1.573	0.009
29	1.318	1.317	0.002
30	1.752	1.744	0.007
31	1.836	1.830	0.006
32	1.507	1.509	0.002
33	1.710	1.703	0.007
34	1.294	1.296	0.002
35	1.397	1.395	0.002
36	1.835	1.832	0.004

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
37	1.700	1.699	0.002
38	1.813	1.809	0.004
39	1.773	1.766	0.007
40	1.386	1.392	0.006
41	1.825	1.817	0.007
42	1.355	1.359	0.004
43	1.380	1.372	0.008
44	1.820	1.811	0.009
45	1.825	1.828	0.003
46	1.298	1.299	0.001
47	1.664	1.663	0.001
48	1.675	1.675	0.001
49	1.523	1.528	0.006
50	1.560	1.552	0.008

Figure 4.36 represents the cell-5 actual and predicted capacity for training and testing split of 70:30% with a total number of instants of 50, using the RF algorithm.

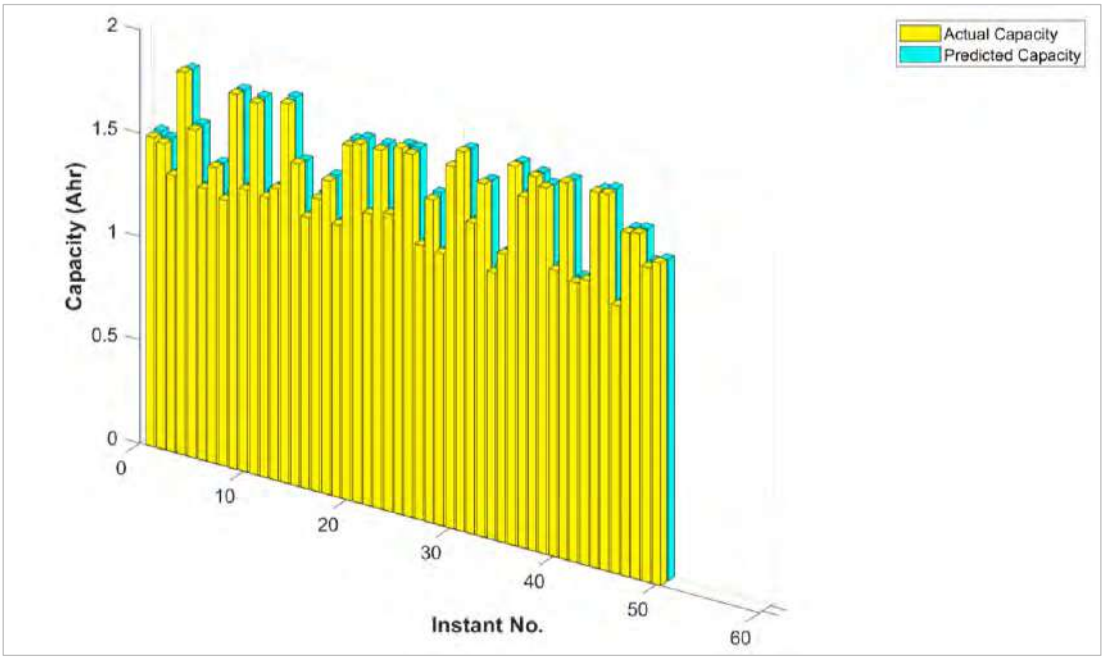


Figure 4.36: Cell-5 actual and predicted capacity for training and testing split of 70:30% using the RF

Figure 4.37 represents the cell-5 actual and predicted capacity for all studied PSs using the RF algorithm.

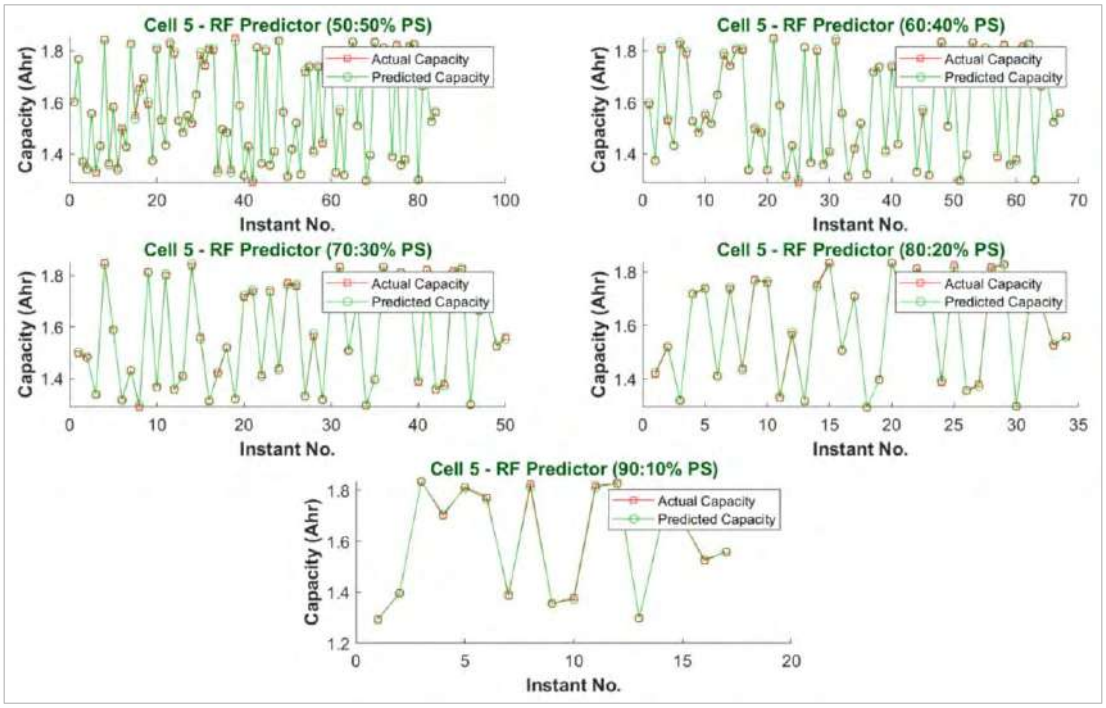


Figure 4.37: Cell-5 actual and predicted capacity for all studied PSs using the RF

Figure 4.38 shows a zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the RF algorithm.

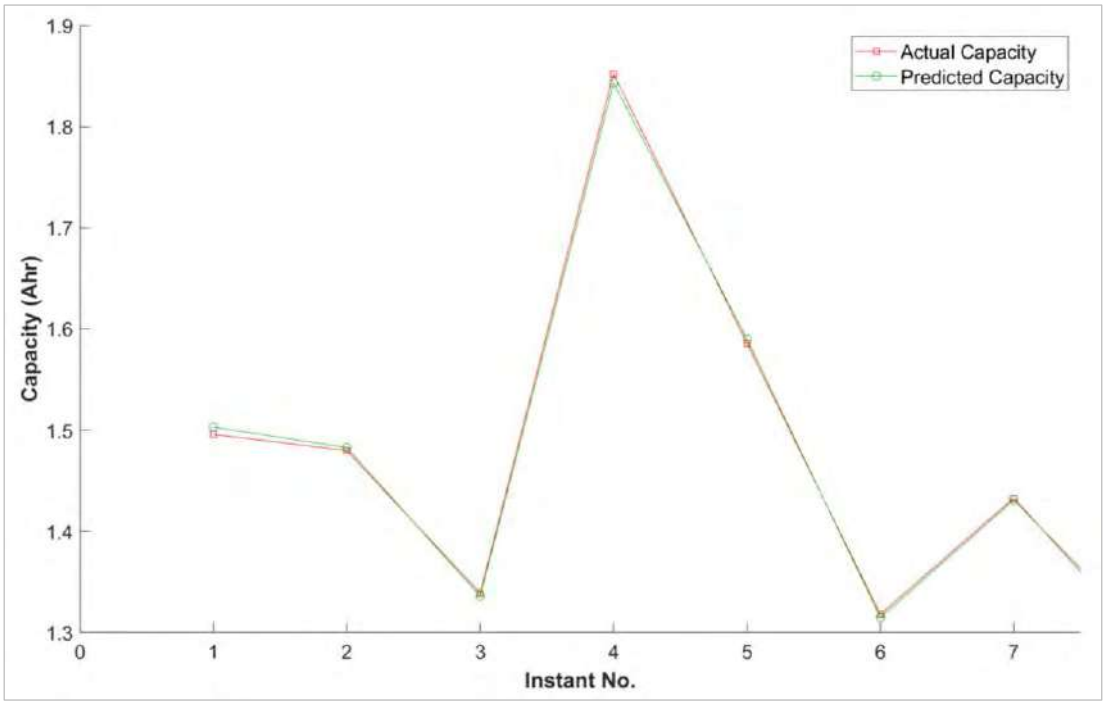


Figure 4.38: Zoom of the cell-5 actual and predicted capacity for the 70:30% PS using the RF

Figure 4.39 till Figure 4.41 represent the cell-5 actual and predicted capacity for The CV of 3-, 5-, and 10-fold, with a total number of instants in each fold of 56, 34 (or 33), and 17 (or 16), respectively, using the RF algorithm.

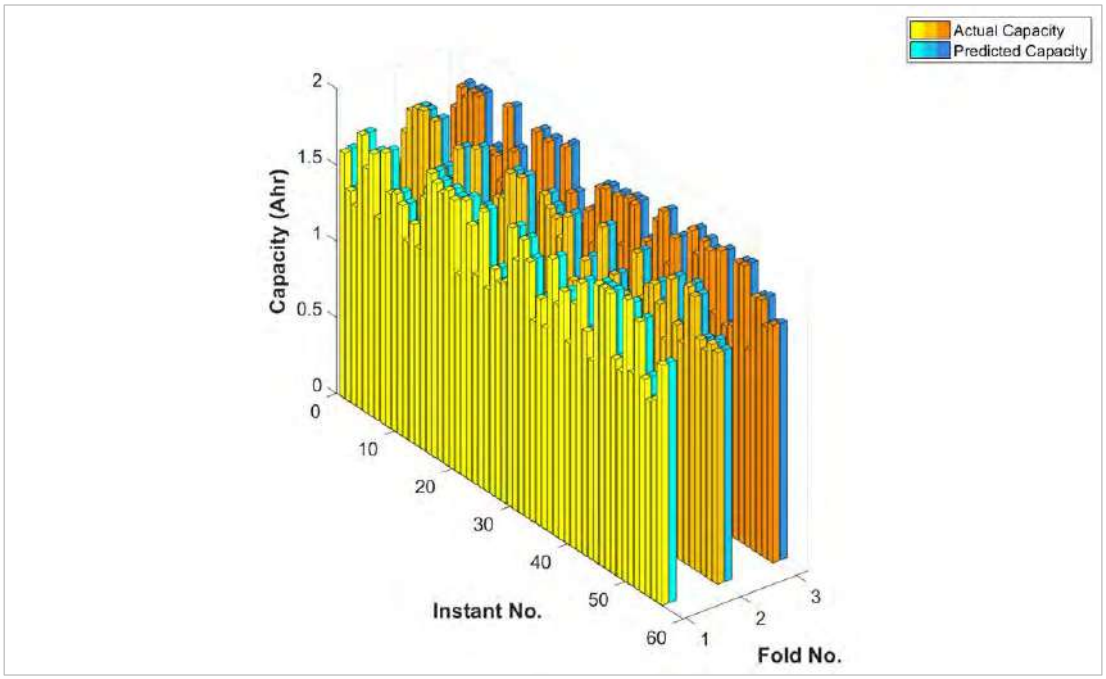


Figure 4.39: Cell-5 actual and predicted capacity for 3-fold CV using the RF

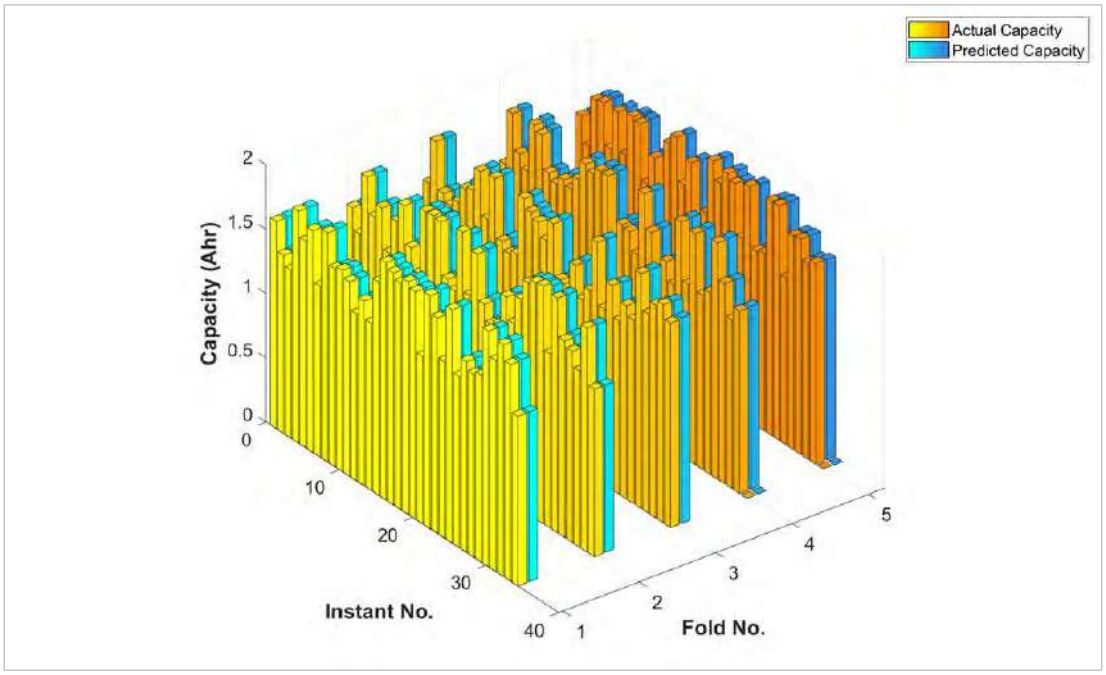


Figure 4.40: Cell-5 actual and predicted capacity for 5-fold CV using the RF

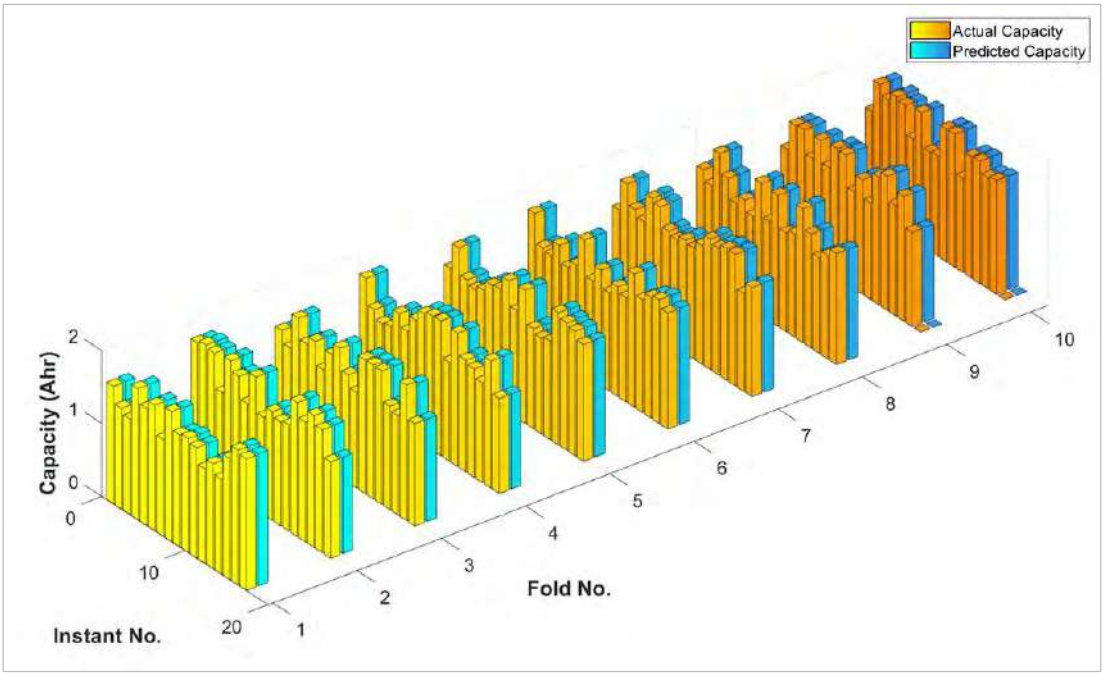


Figure 4.41: Cell-5 actual and predicted capacity for 10-fold CV using the RF

Figure 4.42 represents the cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the RF algorithm.

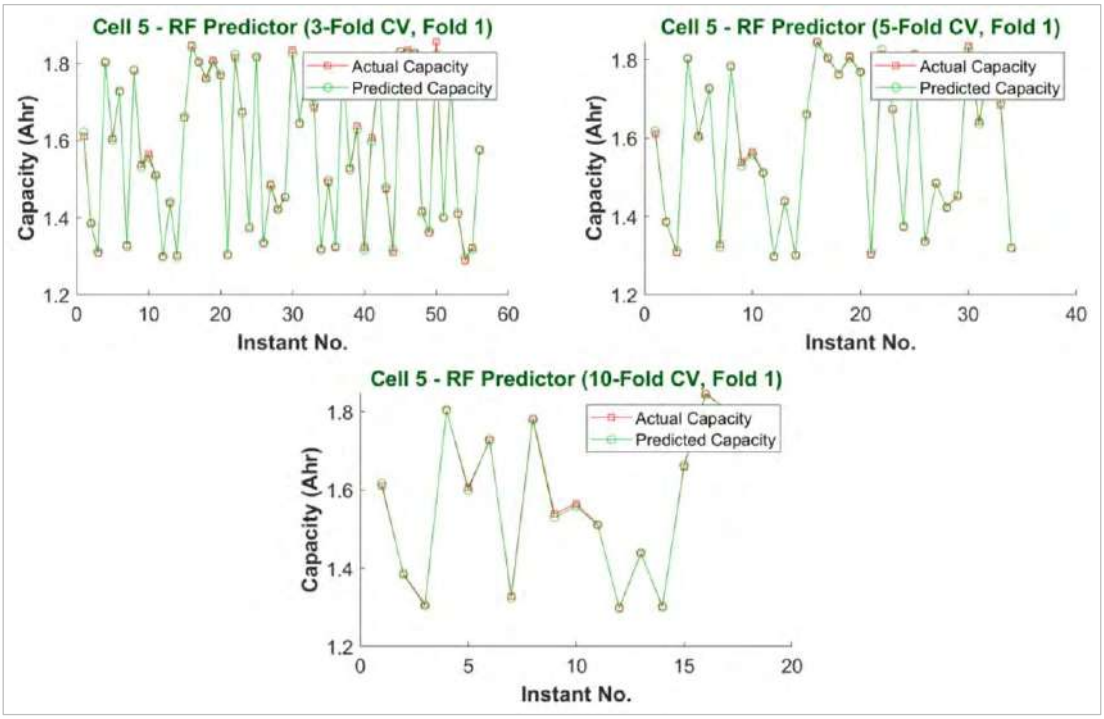


Figure 4.42: Cell-5 actual and predicted capacity for each first fold of the 3-, 5-, and 10-fold CV using the RF

Table A.5 in Appendix-2 shows the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RF algorithm, and the absolute error values, with a total number of instants in this fold of 34.

Figure 4.43 shows a zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RF algorithm.

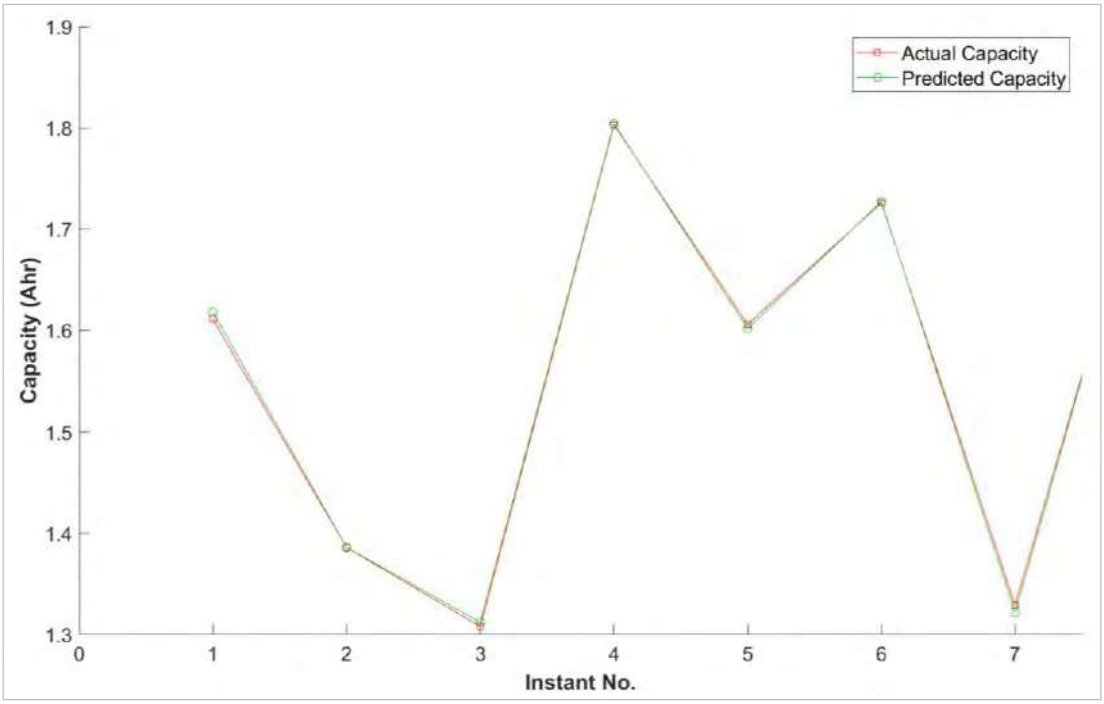


Figure 4.43: Zoom of the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RF

4.3. Prediction Accuracy

The pre-mentioned performance evaluation metrics are calculated for each predictor to assess the suggested solution's accuracy. Table 4.7 presents the results summary for selected methods.

Table 4.7: Performance evaluation metrics comparison for training and testing split of 70:30% and 5-fold CV

Algorithm \ Metric		r	MAE	RMSE	RAE	RRSE
kNN	**70:30%**	0.9997	0.0040	0.0047	2.2223 %	2.4265 %
	5-Fold	0.9995	0.0046	0.0060	2.6890 %	3.1744 %
ANN	**70:30%**	0.9999	0.0034	0.0041	1.8819 %	2.0849 %
	5-Fold	0.9996	0.0038	0.0054	2.2096 %	2.8198 %

Algorithm \ Metric		r	MAE	RMSE	RAE	RRSE
LR	**70:30%**	0.9999	0.0019	0.0023	1.0411 %	1.1606 %
	5-Fold	0.9996	0.0028	0.0051	1.6405 %	2.6656 %
RT	**70:30%**	0.9992	0.0065	0.0077	3.6524 %	3.9630 %
	5-Fold	0.9988	0.0073	0.0094	4.2751 %	4.9558 %
RF	**70:30%**	0.9996	0.0048	0.0054	2.7007 %	2.7748 %
	5-Fold	0.9994	0.0045	0.0064	2.6410 %	3.3917 %

In the context of the PS ratio of 70:30% for the ML algorithms employed, Table 4.7 indicates that both the ANN and LR achieve the maximum correlation coefficient among their studied counterparts in the studied case, with a value of 0.9999. However, the error values show the highest superiority in favor of the LR, which achieves the minimum MAE, RMSE, RAE, and RRSE with values of 0.0019, 0.0023, 1.0411%, 1.1606%, respectively. The ANN attains the second-best performance with the same correlation coefficient as the LR, and values of MAE, RMSE, RAE, and RRSE of 0.0034, 0.0041, 1.8819%, and 2.0849%, respectively. The third-best performance is recorded by the kNN, with a correlation coefficient of 0.9997, and MAE, RMSE, RAE, and RRSE of 0.0040, 0.0047, 2.2223%, and 2.4265%, respectively. After the kNN, the RF comes with a correlation coefficient of 0.9996, and values of MAE, RMSE, RAE, and RRSE of 0.0048, 0.0054, 2.7007%, and 2.7748%, respectively. Lastly, the RT achieves the lowest correlation coefficient among its studied counterparts, with a value of 0.9992, and the highest MAE, RMSE, RAE, and RRSE of 0.0065, 0.0077, 3.6524%, and 3.9630%, respectively. Moreover, the 5-fold CV confirms the 70:30% PS ratio results, with slight differences in the values.

It should be noted that even the lower accuracy values achieved by some of the algorithms used are still considered good values as long as the correlation coefficient is extremely close to 1, and the errors are less than 5%; however, it depends on the application, how much accuracy it needs. In addition to that, these algorithms may be more fit for cases other than the studied case.

Finally, to compare the results of this work with the most important previous studies mentioned in the literature review, we can rely on the RMSE values since it is

the common evaluation metric among all these studies. The study of Zhang et al. [16] achieved an average RMSEs of 0.0335 and 0.0201 at 40 and 70 cycles, while Hu et al. [33] achieved 1.08875, and Wang et al. [34] achieved 0.0082. Fang et al. [39] achieved an RMSE of 0.01156 at 100 cycles, while Li et al. [40] achieved RMSEs of 0.0158 and 0.0130 at 60 and 80 inspection cycles. Shen et al. [41] achieved an RMSE of 3.889. On the other hand, the current work achieved RMSEs for the best algorithm (LR) of 0.0023 and 0.0051 for the PS ratio of 70:30% and the 5-fold CV, respectively. Thus, it excelled the previous counterparts mentioned and achieved RMSE values of less than each of them. That result was expected because the used type of LR algorithm (the least square linear regression) works on squared error reduction.

CHAPTER 5
CONCLUSION AND RECOMMENDATIONS

Reliance on green energy sources has expanded in pursuit of Sustainable Development Goals (SDGs) 7 and 11: "Ensure access to affordable, reliable, sustainable, and modern energy" and "Make cities and human settlements inclusive, safe, resilient and sustainable," respectively. These sources provide an abundant amount of clean energy, but on the other hand, not always available, as many factors affect their continuity and strength of existence; this causes problems in terms of power intermittency and grid instability. Thus, energy storage technologies, which are a key component of any smart grid system and the perfect partner for renewable energy, are being relied on that provide continuous power and contribute to the stability of networks. Among the many of these technologies, Li-ion batteries, which are one of the most important and mature of these technologies, are distinguished by several properties, including but not limited to having light weights, high specific energy density, no memory effect, very high efficiency, and relatively long cycle life. That is in addition to being used in various vital fields, including, for example, hybrid and electric vehicles, drones, and mobile phones.

This thesis focused on upgrading the present Li-ion battery management systems (BMSs), reconstructing their associative data acquisition by employing the event-driven sensing mechanism to reduce overhead usage. Furthermore, it aimed at efficiently predicting the Li-ion battery's remaining useful life (RUL) in pursuit of a safer and more reliable Li-ion battery in the smart grid. It suggested a novel mechanism for event-driven feature extraction to determine the Li-ion battery capacity. Hence, it implicitly determines the RUL of the Li-ion battery so that real-time data compression is achieved. Robust machine learning (ML) classifiers, namely k-Nearest Neighbor (kNN), Artificial Neural Network (ANN), Linear Regression (LR), Random Tree (RT), and Random Forest (RF), are employed to classify the features. They are in the highest accuracy order: LR, ANN, kNN, RF, then RT.

The results show a considerable compression gain of 437.5-fold on average and superior performance for the LR among its studied counterparts. It achieved the minimum MAE, RMSE, RAE, and RRSE with values of 0.0019, 0.0023, 1.0411%, and 1.1606%, respectively, based on the 70:30% PS ratio. Moreover, it achieved the

minimum MAE, RMSE, RAE, and RRSE with values of 0.00289, 0.0051, 1.6405%, and 2.6656%, respectively, based on the 5-fold CV.

Applying the proposed solution on other cells from the intended NASA dataset, other than cell-5, will confirm the robustness of the devised system. Also, extending the work using other ML-based predictors other than those employed is another prospect and future work.

The solution proposed by this work is innovative and is presented in the scientific arena for the first time. Moreover, it is parametrizable and can be utilized in different expected applications such as smart and microgrids, hybrid electric vehicles (HEVs), drones, distributed sensors, satellites, and various renewable energy and power systems applications. Furthermore, the proposed research is well aligned with Saudi Arabia's Vision for 2030. It is a possible candidate to be incorporated into current smart energy systems and integrated into future smart cities like NEOM. Thus, it is recommended for these local and international entities interested in renewable energy and storage technologies in smart grids to employ the proposed solution in designing and advancing power-efficient BMSs for Li-ion batteries. Moreover, the proposed solution has industrial applications; it can be industrialized and commercialized by collaborating with potential industrial partners in Saudi Arabia and other governmental institutions. The devised system can be employed in various applications like integrating renewable energy sources in smart grid, mitigation of power quality issues, effective dimensioning of renewable energy systems, efficient cell-balancing, and energy storage automatic management and maintenance.

REFERENCES

[1] "World Population Prospects 2019," *United Nations, Department of Economic and Social Affairs, Population Division*, Aug. 2019. https://population.un.org/wpp/Download/Standard/Population/ (accessed May 31, 2021).

[2] "Annual Statistics 2020," *The Saudi Central Bank (SAMA)*, May 31, 2021. https://www.sama.gov.sa/en-US/EconomicReports/Pages/YearlyStatistics.aspx (accessed May 31, 2021).

[3] "A Brief on Saudi Arabia's Energy Efficiency Program (SEEP)," *Q. J. Debating Energy Issues Policies*, no. 96, 2014, Accessed: Nov. 12, 2020. [Online]. Available: https://www.oxfordenergy.org/wpcms/wp-content/uploads/2014/06/OEF-96.pdf

[4] K. Malik, S. M. Rahman, A. N. Khondaker, I. R. Abubakar, Y. A. Aina, and M. A. Hasan, "Correction to: Renewable energy utilization to promote sustainability in GCC countries: policies, drivers, and barriers," *Environ. Sci. Pollut. Res.*, vol. 26, no. 30, pp. 31550–31551, Oct. 2019, doi: 10.1007/s11356-019-06138-2.

[5] "Energy & Sustainability," *Vision 2030*. https://www.vision2030.gov.sa/thekingdom/explore/energy/ (accessed May 31, 2021).

[6] "A Global Compact for Sustainable Development – Business and the SDGs: Acting Responsibly and Finding Opportunities | UN Global Compact," *UN Global Compact*, 2015. https://www.unglobalcompact.org/library/2291 (accessed Jan. 24, 2021).

[7] F. Alharbi and D. Csala, "Saudi Arabia's Solar and Wind Energy Penetration: Future Performance and Requirements," *Energies*, vol. 13, no. 3, Art. no. 3, Jan. 2020, doi: 10.3390/en13030588.

[8] A. H. Alattar *et al.*, "Performance Enhancement of Micro Grid System with SMES Storage System Based on Mine Blast Optimization Algorithm," *Energies*, vol. 12, no. 16, Art. no. 16, Jan. 2019, doi: 10.3390/en12163110.

[9] H. Mahmood, T. T. Y. Alkhateeb, and M. Furqan, "Industrialization, urbanization and CO2 emissions in Saudi Arabia: Asymmetry analysis," *Energy Rep.*, vol. 6, pp. 1553–1560, Nov. 2020, doi: 10.1016/j.egyr.2020.06.004.

[10] B. Anthony Jnr., "Integrating Electric Vehicles to Achieve Sustainable Energy as a Service Business Model in Smart Cities," *Front. Sustain. Cities*, vol. 3, p. 45, 2021, doi: 10.3389/frsc.2021.685716.

[11] D. Xu, L. Wang, and J. Yang, "Research on Li-ion Battery Management System," in *2010 International Conference on Electrical and Control Engineering*, Jun. 2010, pp. 4106–4109. doi: 10.1109/iCECE.2010.998.

[12] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 834–854, Oct. 2017, doi: 10.1016/j.rser.2017.05.001.

[13] M. A. Salam and S. A. Khan, "Transition towards sustainable energy production–A review of the progress for solar energy in Saudi Arabia," *Energy Explor. Exploit.*, vol. 36, no. 1, pp. 3–27, 2018.

[14] F. Alfawzan, J. E. Alleman, and C. R. Rehmann, "Wind energy assessment for NEOM city, Saudi Arabia," *Energy Sci. Eng.*, 2019.

[15] O. Hafez, "Time of use prices considering smart meters and their future implementation in Saudi Arabia smart grid," in *2017 Saudi Arabia Smart Grid (SASG)*, 2017, pp. 1–5.

[16] H. Zhang, Q. Miao, X. Zhang, and Z. Liu, “An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction,” *Microelectron. Reliab.*, vol. 81, pp. 288–298, Feb. 2018, doi: 10.1016/j.microrel.2017.12.036.

[17] S. Qaisar, “A Proficient Li-Ion Battery State of Charge Estimation Based on Event-Driven Processing,” *J. Electr. Eng. Technol.*, vol. 15, Jun. 2020, doi: 10.1007/s42835-020-00458-x.

[18] R. C. Massé, E. Uchaker, and G. Cao, “Beyond Li-ion: electrode materials for sodium- and magnesium-ion batteries,” *Sci. China Mater.*, vol. 58, no. 9, pp. 715–766, Sep. 2015, doi: 10.1007/s40843-015-0084-8.

[19] X. Bian, L. Liu, J. Yan, Z. Zou, and R. Zhao, “An open circuit voltage-based model for state-of-health estimation of lithium-ion batteries: Model development and validation,” *J. Power Sources*, vol. 448, p. 227401, Feb. 2020, doi: 10.1016/j.jpowsour.2019.227401.

[20] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, “Remaining useful life prediction of lithium-ion battery with unscented particle filter technique,” *Microelectron. Reliab.*, vol. 53, no. 6, pp. 805–810, Jun. 2013, doi: 10.1016/j.microrel.2012.12.004.

[21] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, “Critical review of state of health estimation methods of Li-ion batteries for real applications,” *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016, doi: 10.1016/j.rser.2015.11.042.

[22] B. Long, W. Xian, L. Jiang, and Z. Liu, “An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries,” *Microelectron. Reliab.*, vol. 53, no. 6, pp. 821–831, Jun. 2013, doi: 10.1016/j.microrel.2013.01.006.

[23] G. Berckmans, M. Messagie, J. Smekens, N. Omar, L. Vanhaverbeke, and J. Van Mierlo, “Cost Projection of State of the Art Lithium-Ion Batteries for Electric Vehicles Up to 2030,” *Energies*, vol. 10, no. 9, Art. no. 9, Sep. 2017, doi: 10.3390/en10091314.

[24] S. S. Y. Ng, Y. Xing, and K. L. Tsui, “A naive Bayes model for robust remaining useful life prediction of lithium-ion battery,” *Appl. Energy*, vol. 118, pp. 114–123, Apr. 2014, doi: 10.1016/j.apenergy.2013.12.020.

[25] X. Han *et al.*, “A review on the key issues of the lithium ion battery degradation among the whole life cycle,” *eTransportation*, vol. 1, p. 100005, Aug. 2019, doi: 10.1016/j.etran.2019.100005.

[26] “Lithium Cell and Battery Standard.” University of Waterloo - Safety Office, Jul. 02, 2019. Accessed: Jun. 02, 2021. [Online]. Available: https://uwaterloo.ca/safety-office/laboratory-safety/batteries

[27] M. A. Patil *et al.*, “A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation,” *Appl. Energy*, vol. 159, pp. 285–297, Dec. 2015, doi: 10.1016/j.apenergy.2015.08.119.

[28] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, and L. Zhang, “Remaining Useful Life Prediction for Lithium-Ion Battery: A Deep Learning Approach,” *IEEE Access*, vol. 6, pp. 50587–50598, 2018, doi: 10.1109/ACCESS.2018.2858856.

[29] B. Kumar, N. Khare, and P. K. Chaturvedi, “Advanced battery management system using MATLAB/Simulink,” in *2015 IEEE International Telecommunications Energy Conference (INTELEC)*, Oct. 2015, pp. 1–6. doi: 10.1109/INTLEC.2015.7572447.

[30] M. Kokila, P. Manimekalai, and V. Indragandhi, “Design and development of battery management system (BMS) using hybrid multilevel converter,” *Int. J. Ambient Energy*, vol. 41, no. 7, pp. 729–737, Jun. 2020, doi: 10.1080/01430750.2018.1492440.

[31] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, “Critical Review on the Battery State of Charge Estimation Methods for Electric Vehicles,” *IEEE Access*, vol. 6, pp. 1832–1843, 2018, doi: 10.1109/ACCESS.2017.2780258.

[32] Y. Li *et al.*, “A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter,” *J. Power Sources*, vol. 373, pp. 40–53, Jan. 2018, doi: 10.1016/j.jpowsour.2017.10.092.

[33] C. Hu, G. Jain, P. Zhang, C. Schmidt, P. Gomadam, and T. Gorka, “Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery,” *Appl. Energy*, vol. 129, pp. 49–55, Sep. 2014, doi: 10.1016/j.apenergy.2014.04.077.

[34] D. Wang, Q. Miao, and M. Pecht, “Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model,” *J. Power Sources*, vol. 239, pp. 253–264, Oct. 2013, doi: 10.1016/j.jpowsour.2013.03.129.

[35] D. Z. Li, W. Wang, and F. Ismail, “A Mutated Particle Filter Technique for System State Estimation and Battery Life Prediction,” *IEEE Trans. Instrum. Meas.*, vol. 63, no. 8, pp. 2034–2043, Aug. 2014, doi: 10.1109/TIM.2014.2303534.

[36] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, “Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695–5705, Jul. 2018, doi: 10.1109/TVT.2018.2805189.

[37] Y. Chen, Q. Miao, B. Zheng, S. Wu, and M. Pecht, “Quantitative Analysis of Lithium-Ion Battery Capacity Prediction via Adaptive Bathtub-Shaped Function,” *Energies*, vol. 6, no. 6, Art. no. 6, Jun. 2013, doi: 10.3390/en6063082.

[38] S. Tang, C. Yu, X. Wang, X. Guo, and X. Si, “Remaining Useful Life Prediction of Lithium-Ion Batteries Based on the Wiener Process with Measurement Error,” *Energies*, vol. 7, no. 2, Art. no. 2, Feb. 2014, doi: 10.3390/en7020520.

[39] X. Zheng and H. Fang, “An integrated unscented kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction,” *Reliab. Eng. Syst. Saf.*, vol. 144, pp. 74–82, Dec. 2015, doi: 10.1016/j.ress.2015.07.013.

[40] L. Li, P. Wang, K.-H. Chao, Y. Zhou, and Y. Xie, “Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Gaussian Processes Mixture,” *PloS One*, vol. 11, p. e0163004, Sep. 2016, doi: 10.1371/journal.pone.0163004.

[41] D. Shen, L. Wu, G. Kang, Y. Guan, and Z. Peng, “A novel online method for predicting the remaining useful life of lithium-ion batteries considering random variable discharge current,” *Energy*, vol. 218, p. 119490, Mar. 2021, doi: 10.1016/j.energy.2020.119490.

[42] B. Saha and K. Goebel, “Prognostics Center of Excellence - Battery Data Set,” *NASA Ames Research Center*, 2008. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/ (accessed Jan. 28, 2021).

[43] B. Saha and K. Goebel, “Battery Data Set,” *NASA Ames Research Center*, 2008. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/ (accessed Dec. 20, 2020).

[44] B. A. Forouzan and S. C. Fegan, *Data Communications and Networking*, 4th ed. Huga Media, 2007.

[45] W.-K. Ling, “1 - INTRODUCTION,” in *Nonlinear Digital Filters*, W.-K. Ling, Ed. Oxford: Academic Press, 2007, pp. 1–7. doi: 10.1016/B978-012372536-3/50001-6.

[46] S. M. Qaisar, “A Computationally Efficient EEG Signals Segmentation and De-noising Based on an Adaptive Rate Acquisition and Processing,” in *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*, 2018, pp. 182–186.

[47] S. M. Qaisar, L. Fesquet, and M. Renaudin, “Adaptive rate filtering a computationally efficient signal processing approach,” *Signal Process.*, vol. 94, pp. 620–630, 2014.

[48] S. M. Qaisar, D. Dallet, S. Benjamin, P. Desprez, and R. Yahiaoui, “Power efficient analog to digital conversion for the Li-ion battery voltage monitoring and measurement,” in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 1522–1525.

[49] S. M. Qaisar, R. Yahiaoui, and D. Dominique, “A smart power management system monitoring and measurement approach based on a signal driven data acquisition,” in *2015 Saudi Arabia Smart Grid (SASG)*, 2015, pp. 1–4.

[50] S. Mian Qaisar, “Event-Driven Coulomb Counting for Effective Online Approximation of Li-Ion Battery State of Charge,” *Energies*, vol. 13, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/en13215600.

[51] E. Allier, G. Sicard, L. Fesquet, and M. Renaudin, “A new class of asynchronous A/D converters based on time quantization,” in *Ninth International Symposium on Asynchronous Circuits and Systems, 2003. Proceedings.*, May 2003, pp. 196–205. doi: 10.1109/ASYNC.2003.1199179.

[52] V. K. Ingle and J. G. Proakis, *Digital signal processing using matlab: a problem solving companion*. Cengage Learning, 2016.

[53] S. Qaisar and F. Alsharif, “Event-Driven System for Proficient Load Recognition by Interpreting the Smart Meter Data,” *Procedia Comput. Sci.*, vol. 168, pp. 210–216, Jan. 2020, doi: 10.1016/j.procs.2020.02.267.

[54] M. Pandey and S. S. Rautaray, Eds., *Machine Learning: Theoretical Foundations and Practical Applications*, vol. 87. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-33-6518-6.

[55] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[56] L. Mohan, S. Jain, P. Suyal, and A. Kumar, “Data mining Classification Techniques for Intrusion Detection System,” in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2020, pp. 351–355. doi: 10.1109/CICN49253.2020.9242642.

[57] Y. Zhao and Y. Zhang, “Comparison of decision tree methods for finding active objects,” *Adv. Space Res.*, vol. 41, no. 12, pp. 1955–1959, Jan. 2008, doi: 10.1016/j.asr.2007.07.020.

[58] Y.-S. Chen, A. K. Sangaiah, S.-F. Chen, and H.-C. Huang, “Applied Identification of Industry Data Science Using an Advanced Multi-Componential Discretization Model,” *Symmetry*, vol. 12, no. 10, Art. no. 10, Oct. 2020, doi: 10.3390/sym12101620.

[59] J. Brownlee, “Train-Test Split for Evaluating Machine Learning Algorithms,” *Machine Learning Mastery*, Jul. 23, 2020. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/ (accessed May 25, 2021).

[60] D. Berrar, “Cross-Validation,” in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. Oxford: Academic Press, 2019, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[61] J. Mun, “Chapter 3 - A Primer on Quantitative Risk Analysis,” in *Multi-Asset Risk Modeling*, M. Glantz and R. Kissell, Eds. San Diego: Academic Press, 2014, pp. 63–118. doi: 10.1016/B978-0-12-401690-3.00003-2.

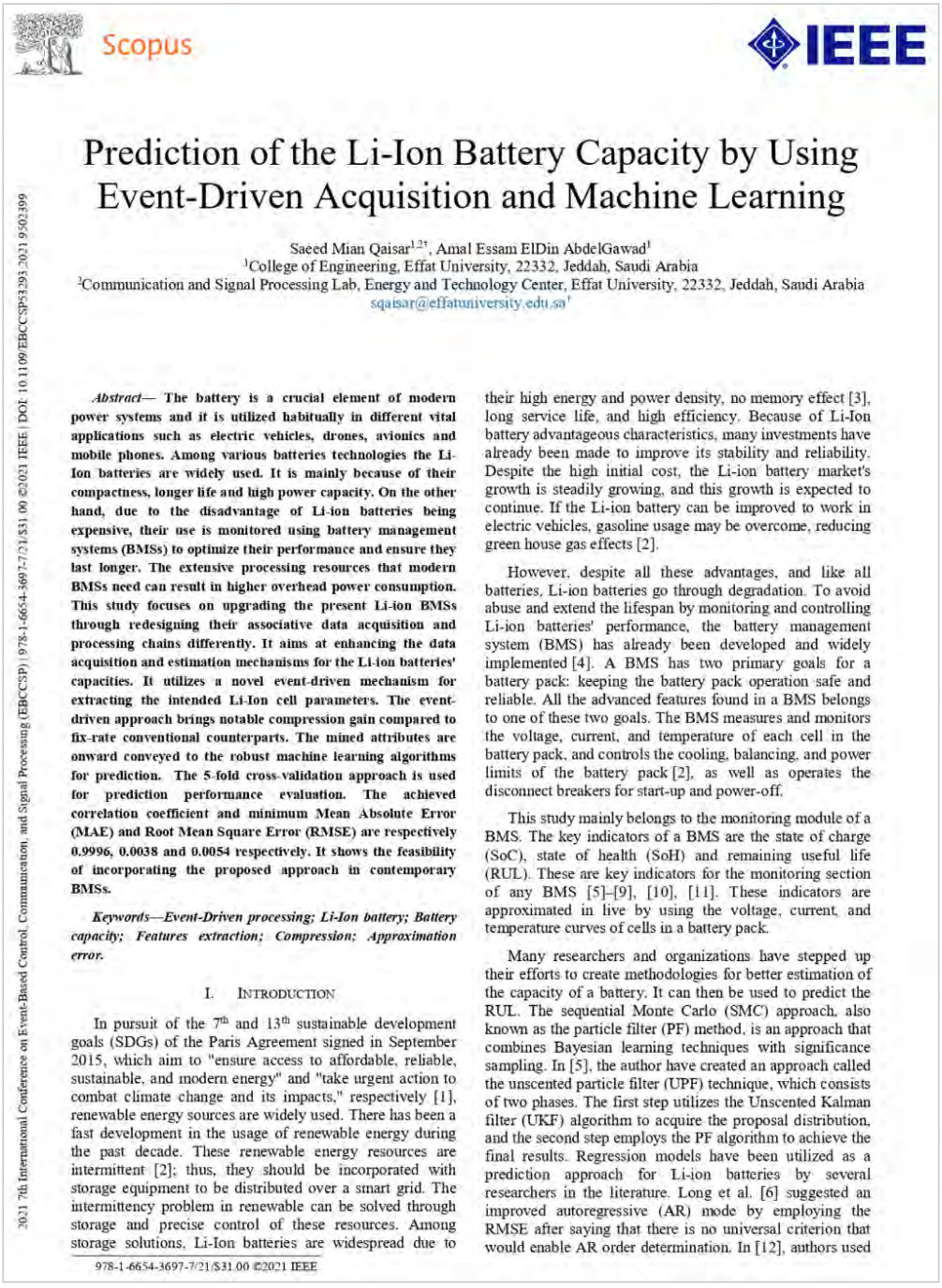
[62] L. Zavarella, “How to Better Evaluate the Goodness-of-Fit of Regressions,” *Medium*, Feb. 05, 2019. https://medium.com/microsoftazure/how-to-better-evaluate-the-goodness-of-fit-of-regressions-990dbf1c0091 (accessed Feb. 24, 2021).

[63] M. Lucu, E. Martinez-Laserna, I. Gandiaga, and H. Camblong, “A critical review on self-adaptive Li-ion battery ageing models,” *J. Power Sources*, vol. 401, pp. 85–101, Oct. 2018, doi: 10.1016/j.jpowsour.2018.08.064.

APPENDICES

Appendix-1: Published Papers

A1.1. A paper that got submitted and presented at the 7th International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP) entitled "**Prediction of the Li-Ion Battery Capacity by Using Event-Driven Acquisition and Machine Learning**." The conference was held virtual between the 22nd and 25th of June 2021. The paper has been published in IEEE Xplore under the DOI: 10.1109/EBCCSP53293.2021.9502399 and indexed in Scopus.



Scopus

IEEE

Prediction of the Li-Ion Battery Capacity by Using Event-Driven Acquisition and Machine Learning

Saeed Mian Qaisar[1,2†], Amal Essami ElDin AbdelGawad[1]
[1]College of Engineering, Effat University, 22332, Jeddah, Saudi Arabia
[2]Communication and Signal Processing Lab, Energy and Technology Center, Effat University, 22332, Jeddah, Saudi Arabia
sqaisar@effatuniversity.edu.sa[†]

2021 7th International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP) | 978-1-6654-3697-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EBCCSP53293.2021.9502399

***Abstract*— The battery is a crucial element of modern power systems and it is utilized habitually in different vital applications such as electric vehicles, drones, avionics and mobile phones. Among various batteries technologies the Li-Ion batteries are widely used. It is mainly because of their compactness, longer life and high power capacity. On the other hand, due to the disadvantage of Li-ion batteries being expensive, their use is monitored using battery management systems (BMSs) to optimize their performance and ensure they last longer. The extensive processing resources that modern BMSs need can result in higher overhead power consumption. This study focuses on upgrading the present Li-ion BMSs through redesigning their associative data acquisition and processing chains differently. It aims at enhancing the data acquisition and estimation mechanisms for the Li-ion batteries' capacities. It utilizes a novel event-driven mechanism for extracting the intended Li-Ion cell parameters. The event-driven approach brings notable compression gain compared to fix-rate conventional counterparts. The mined attributes are onward conveyed to the robust machine learning algorithms for prediction. The 5-fold cross-validation approach is used for prediction performance evaluation. The achieved correlation coefficient and minimum Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are respectively 0.9996, 0.0038 and 0.0054 respectively. It shows the feasibility of incorporating the proposed approach in contemporary BMSs.**

Keywords—Event-Driven processing; Li-Ion battery; Battery capacity; Features extraction; Compression; Approximation error.

I. INTRODUCTION

In pursuit of the 7th and 13th sustainable development goals (SDGs) of the Paris Agreement signed in September 2015, which aim to "ensure access to affordable, reliable, sustainable, and modern energy" and "take urgent action to combat climate change and its impacts," respectively [1], renewable energy sources are widely used. There has been a fast development in the usage of renewable energy during the past decade. These renewable energy resources are intermittent [2]; thus, they should be incorporated with storage equipment to be distributed over a smart grid. The intermittency problem in renewable can be solved through storage and precise control of these resources. Among storage solutions, Li-Ion batteries are widespread due to their high energy and power density, no memory effect [3], long service life, and high efficiency. Because of Li-Ion battery advantageous characteristics, many investments have already been made to improve its stability and reliability. Despite the high initial cost, the Li-ion battery market's growth is steadily growing, and this growth is expected to continue. If the Li-ion battery can be improved to work in electric vehicles, gasoline usage may be overcome, reducing green house gas effects [2].

However, despite all these advantages, and like all batteries, Li-ion batteries go through degradation. To avoid abuse and extend the lifespan by monitoring and controlling Li-ion batteries' performance, the battery management system (BMS) has already been developed and widely implemented [4]. A BMS has two primary goals for a battery pack: keeping the battery pack operation safe and reliable. All the advanced features found in a BMS belongs to one of these two goals. The BMS measures and monitors the voltage, current, and temperature of each cell in the battery pack, and controls the cooling, balancing, and power limits of the battery pack [2], as well as operates the disconnect breakers for start-up and power-off.

This study mainly belongs to the monitoring module of a BMS. The key indicators of a BMS are the state of charge (SoC), state of health (SoH) and remaining useful life (RUL). These are key indicators for the monitoring section of any BMS [5]–[9], [10], [11]. These indicators are approximated in live by using the voltage, current, and temperature curves of cells in a battery pack.

Many researchers and organizations have stepped up their efforts to create methodologies for better estimation of the capacity of a battery. It can then be used to predict the RUL. The sequential Monte Carlo (SMC) approach, also known as the particle filter (PF) method, is an approach that combines Bayesian learning techniques with significance sampling. In [5], the author have created an approach called the unscented particle filter (UPF) technique, which consists of two phases. The first step utilizes the Unscented Kalman filter (UKF) algorithm to acquire the proposal distribution, and the second step employs the PF algorithm to achieve the final results. Regression models have been utilized as a prediction approach for Li-ion batteries by several researchers in the literature. Long et al. [6] suggested an improved autoregressive (AR) mode by employing the RMSE after saying that there is no universal criterion that would enable AR order determination. In [12], authors used

978-1-6654-3697-7/21/$31.00 ©2021 IEEE

A1.2. A paper that got submitted and presented at the EASTON 100 – 5th International Conference on Computing Sciences (ICCS 2021) entitled "**Event-Driven Acquisition and Machine-Learning-Based Efficient Prediction of the Li-Ion Battery Capacity**." The conference was held virtual and organized by the Advanced Computing Research Society (ACRS), Chennai, India, between the 30th of April and the 1st of May 2021. The paper has been published in Springer Nature Computer Science Journal, 3, 15 (2022) on the 26th of October 2021, under the DOI: 10.1007/s42979-021-00905-0, and indexed in Scopus.

SN Computer Science (2022) 3:15
https://doi.org/10.1007/s42979-021-00905-0

Scopus

SN

ORIGINAL RESEARCH

Event-Driven Acquisition and Machine-Learning-Based Efficient Prediction of the Li-Ion Battery Capacity

Saeed Mian Qaisar[1,2] · Amal Essam ElDin AbdelGawad[1] · Kathiravan Srinivasan[3]

Received: 21 August 2021 / Accepted: 22 September 2021

Abstract
In almost all contemporary power systems, the battery is an elementary component, and it is routinely used in a variety of critical applications such as drones, avionics, and cell phones. Due to their superior characteristics compared to the concurrent technologies, Li-ion batteries are widely utilized. Since batteries are costly, their usage is closely monitored by battery management systems (BMSs). It ensures that batteries survive and serve longer. Modern BMSs' are complex and sophisticated and can deal with hundreds of cells in a battery pack. It results in an increased processing resources requirement and can cause an overhead power consumption. The aim of this work is to improve current BMSs by redesigning their associative processing chain. It focuses on improving data collection, processing and prediction processes for Li-ion battery cell capacities. To prevent the processing of a large amount of unnecessary data, the classical sensing approach that is fix-rate is avoided and replaced by event-driven sensing (EDS) mechanism to digitize battery cell parameters such as voltages, currents, and temperatures in a way that allows for real-time data compressing. A new approach is proposed for event-driven feature extraction. The robust machine-learning algorithms are employed for processing the extracted features and to predict the capacity of considered battery cell. Results show a considerable compression gain with a correlation coefficient of 0.999 and the relative absolute error (RAE) and root relative squared error (RRSE) of 1.88% and 2.08%, respectively.

Keywords Event-driven processing · Li-ion battery · Remaining useful life · Battery capacity · Features extraction · Computational complexity

Introduction

The Paris Agreement, which was signed in September 2015, aims to make energy more affordable and sustainable. Its goal is to make cities and human settlements safer and long-lasting [1]. After the percentage of greenhouse gases (GHG) emissions in the Earth's atmosphere rose to new record levels in 2019, it became the second warmest year on record and the warmest year in the last decade (2010–2019). In the first quarter of 2020, demand for carbon fuels had recorded its most significant decline, as the climate improved, and global CO_2 emissions fell twice as much as the previous total declines since the end of World War II in 1945. Despite that, it was only a temporary improvement in the climate, and once the global economy begins to recover from the Covid-19 pandemic, that escalation in GHG levels will return to the increase [2, 3]. The planet is making progress in achieving the 7th SDG, as energy is more sustainable and available, renewable energy efficiency has improved, its use expanded, and its economic performance in the electricity sector has improved [4]. However, the fact that these renewable energy sources, such as wind and solar, are intermittent. As a result, they should be combined with energy storage modules. That is why the importance of renewable power storage has increased, as it provides a practical solution for network control and power stability-enhancing [5].

This article is part of the topical collection "Intelligent Systems" guest edited by Geetha Ganesan, Lalit Garg, Renu Dhir, Vijay Kumar and Manik Sharma.

✉ Saeed Mian Qaisar
sqaisar@effatuniversity.edu.sa

1 College of Engineering, Effat University, Jeddah 22332, Saudi Arabia

2 Communication and Signal Processing Lab, Energy and Technology Research Center, Effat University, Jeddah 22332, Saudi Arabia

3 School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, India

Published online: 26 October 2021

Appendix-2: Some Tabular Results

Table A.1 till Table A.5 show the cell-5 actual and predicted capacity for fold-1 of the 5-fold CV and the absolute error values, with a total number of instants in this fold of 34, using the kNN, ANN, LR, RT, and RF algorithms, respectively.

Table A.1: Cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the kNN

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.611	1.614	0.003
2	1.386	1.379	0.007
3	1.308	1.310	0.002
4	1.803	1.805	0.002
5	1.606	1.606	0.001
6	1.726	1.733	0.006
7	1.329	1.329	0.000
8	1.783	1.775	0.007
9	1.538	1.533	0.005
10	1.565	1.556	0.009
11	1.512	1.517	0.006
12	1.298	1.296	0.002
13	1.438	1.449	0.011
14	1.303	1.296	0.008
15	1.659	1.672	0.013
16	1.847	1.834	0.013
17	1.803	1.809	0.006
18	1.763	1.765	0.002
19	1.809	1.803	0.007
20	1.768	1.769	0.001
21	1.303	1.307	0.004
22	1.814	1.822	0.008
23	1.675	1.672	0.003
24	1.375	1.374	0.001
25	1.814	1.814	0.000

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
26	1.334	1.338	0.004
27	1.486	1.484	0.001
28	1.423	1.428	0.005
29	1.454	1.449	0.005
30	1.835	1.828	0.007
31	1.643	1.638	0.005
32	1.803	1.806	0.003
33	1.685	1.694	0.009
34	1.319	1.320	0.001

Table A.2: Cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the ANN

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.611	1.613	0.002
2	1.386	1.386	0.001
3	1.308	1.308	0.000
4	1.803	1.803	0.000
5	1.606	1.585	0.021
6	1.726	1.725	0.001
7	1.329	1.327	0.002
8	1.783	1.783	0.000
9	1.538	1.535	0.003
10	1.565	1.565	0.000
11	1.512	1.511	0.001
12	1.298	1.298	0.000
13	1.438	1.438	0.000
14	1.303	1.301	0.003
15	1.659	1.661	0.002
16	1.847	1.844	0.003
17	1.803	1.810	0.008

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
18	1.763	1.761	0.002
19	1.809	1.806	0.003
20	1.768	1.768	0.001
21	1.303	1.305	0.002
22	1.814	1.821	0.007
23	1.675	1.674	0.001
24	1.375	1.379	0.004
25	1.814	1.814	0.000
26	1.334	1.334	0.000
27	1.486	1.487	0.001
28	1.423	1.423	0.001
29	1.454	1.454	0.000
30	1.835	1.830	0.005
31	1.643	1.646	0.003
32	1.803	1.802	0.001
33	1.685	1.686	0.001
34	1.319	1.320	0.001

Table A.3: Cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the LR

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.611	1.613	0.002
2	1.386	1.385	0.001
3	1.308	1.309	0.001
4	1.803	1.802	0.001
5	1.606	1.600	0.006
6	1.726	1.726	0.000
7	1.329	1.327	0.002
8	1.783	1.784	0.001
9	1.538	1.536	0.003

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
10	1.565	1.565	0.000
11	1.512	1.511	0.001
12	1.298	1.299	0.001
13	1.438	1.440	0.001
14	1.303	1.301	0.002
15	1.659	1.661	0.002
16	1.847	1.847	0.000
17	1.803	1.810	0.007
18	1.763	1.762	0.001
19	1.809	1.810	0.000
20	1.768	1.769	0.001
21	1.303	1.305	0.002
22	1.814	1.820	0.006
23	1.675	1.674	0.001
24	1.375	1.378	0.002
25	1.814	1.813	0.001
26	1.334	1.334	0.000
27	1.486	1.488	0.002
28	1.423	1.424	0.001
29	1.454	1.455	0.001
30	1.835	1.830	0.005
31	1.643	1.645	0.002
32	1.803	1.800	0.002
33	1.685	1.685	0.000
34	1.319	1.319	0.001

Table A.4: Cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RT

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.611	1.630	0.019

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
2	1.386	1.383	0.003
3	1.308	1.320	0.012
4	1.803	1.803	0.001
5	1.606	1.584	0.022
6	1.726	1.711	0.016
7	1.329	1.320	0.009
8	1.783	1.769	0.014
9	1.538	1.525	0.014
10	1.565	1.556	0.009
11	1.512	1.501	0.011
12	1.298	1.301	0.003
13	1.438	1.453	0.014
14	1.303	1.301	0.003
15	1.659	1.669	0.010
16	1.847	1.851	0.004
17	1.803	1.803	0.000
18	1.763	1.769	0.006
19	1.809	1.803	0.007
20	1.768	1.769	0.001
21	1.303	1.301	0.002
22	1.814	1.830	0.016
23	1.675	1.669	0.006
24	1.375	1.383	0.008
25	1.814	1.815	0.001
26	1.334	1.340	0.006
27	1.486	1.483	0.003
28	1.423	1.433	0.011
29	1.454	1.453	0.001
30	1.835	1.830	0.006
31	1.643	1.630	0.013

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
32	1.803	1.803	0.000
33	1.685	1.697	0.013
34	1.319	1.320	0.001

Table A.5: Cell-5 actual and predicted capacity for fold-1 of the 5-fold CV using the RF

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
1	1.611	1.618	0.007
2	1.386	1.386	0.001
3	1.308	1.312	0.004
4	1.803	1.804	0.001
5	1.606	1.602	0.004
6	1.726	1.727	0.001
7	1.329	1.322	0.007
8	1.783	1.785	0.002
9	1.538	1.529	0.009
10	1.565	1.558	0.007
11	1.512	1.512	0.000
12	1.298	1.297	0.000
13	1.438	1.440	0.001
14	1.303	1.300	0.004
15	1.659	1.661	0.002
16	1.847	1.844	0.003
17	1.803	1.805	0.002
18	1.763	1.763	0.001
19	1.809	1.805	0.004
20	1.768	1.769	0.001
21	1.303	1.307	0.004
22	1.814	1.826	0.012
23	1.675	1.673	0.002

Instant No.	Actual (Ahr)	Predicted (Ahr)	$\|Error\|$
24	1.375	1.374	0.002
25	1.814	1.814	0.001
26	1.334	1.337	0.003
27	1.486	1.485	0.001
28	1.423	1.423	0.000
29	1.454	1.452	0.003
30	1.835	1.829	0.006
31	1.643	1.637	0.005
32	1.803	1.804	0.001
33	1.685	1.692	0.007
34	1.319	1.320	0.001