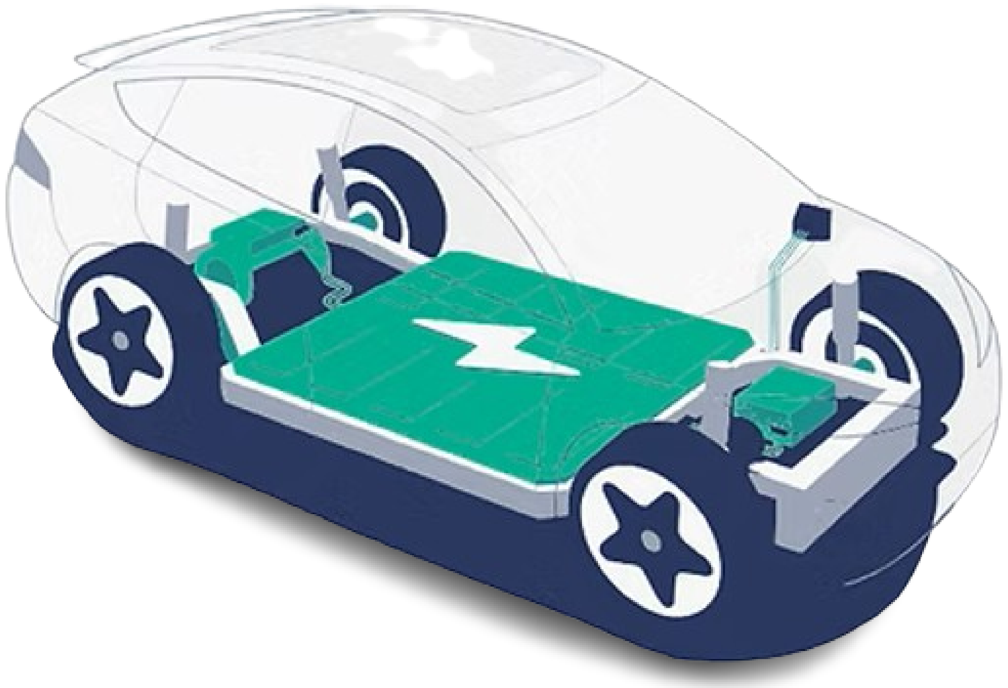
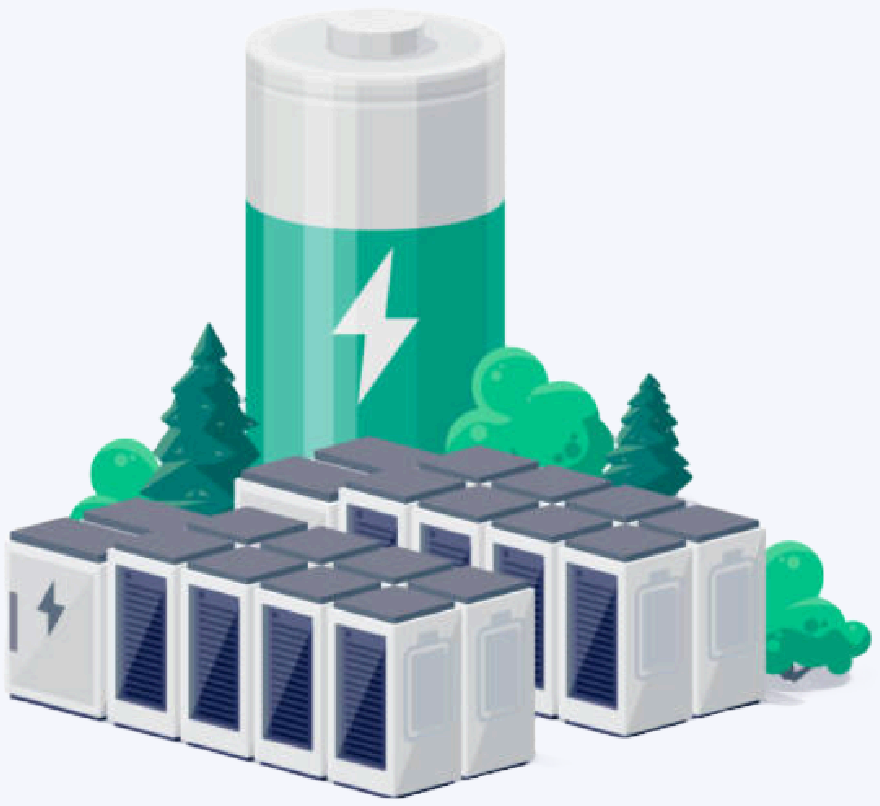
Data-driven prediction of battery cycle life before capacity degradation



Department of Chemical Engineering and Technology
Indian Institute of Technology (BHU), Varanasi

SIGNIFICANCE

- Battery degradation takes hundreds of cycles to appear, making traditional testing slow and expensive.
- Battery life is hard to predict early, as even similar cells often behave very differently.
- Machine learning enables early prediction by detecting hidden degradation patterns from initial cycles.
- Early prediction unlocks faster cell development, quality grading, and better life expectancy estimates for manufacturers and end users



How battery Degrades (Literature review)

Lithium-ion batteries degrade over time and usage due to a variety of intertwined physical and chemical processes. Major degradation mechanisms include:

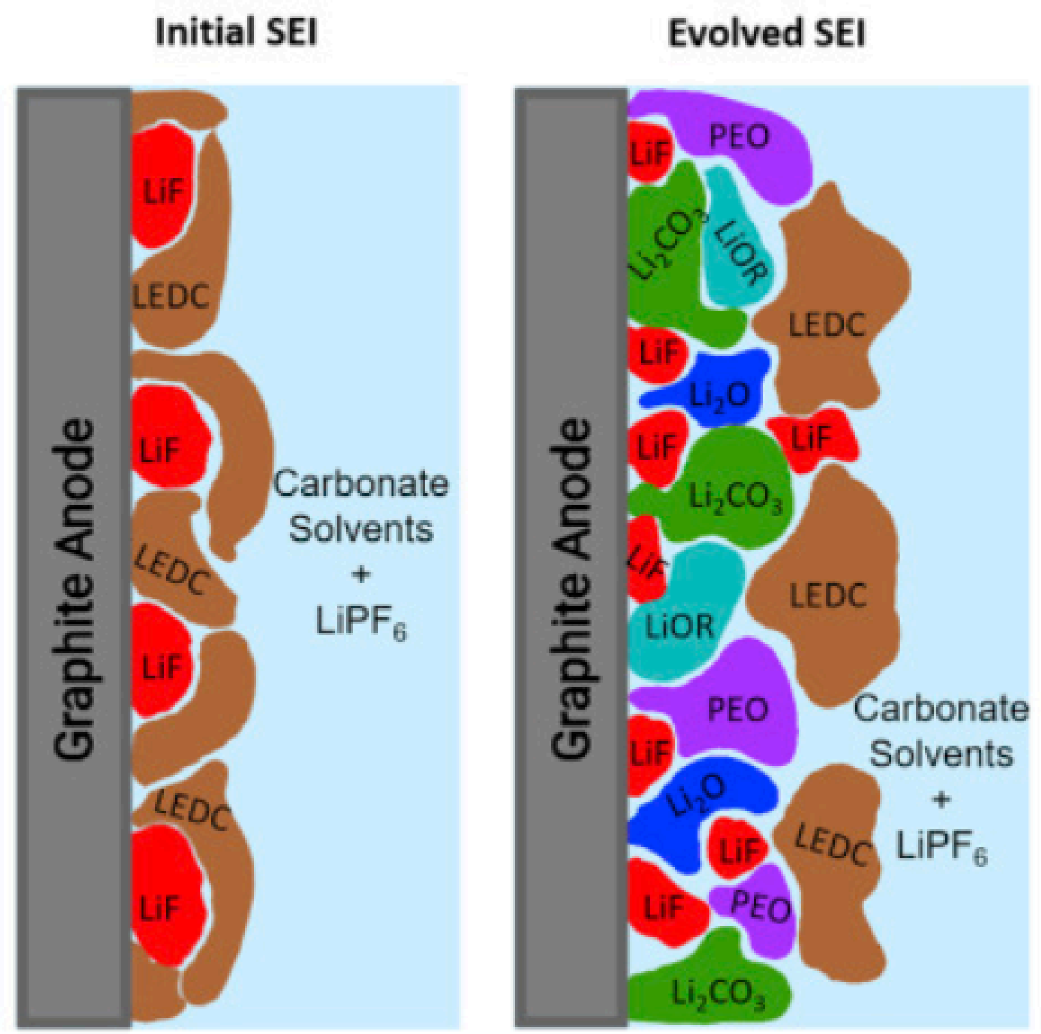
- Solid Electrolyte Interphase (SEI) growth
- Lithium plating
- Positive electrode (cathode) structural changes and decomposition
- Particle fracture (both electrodes)

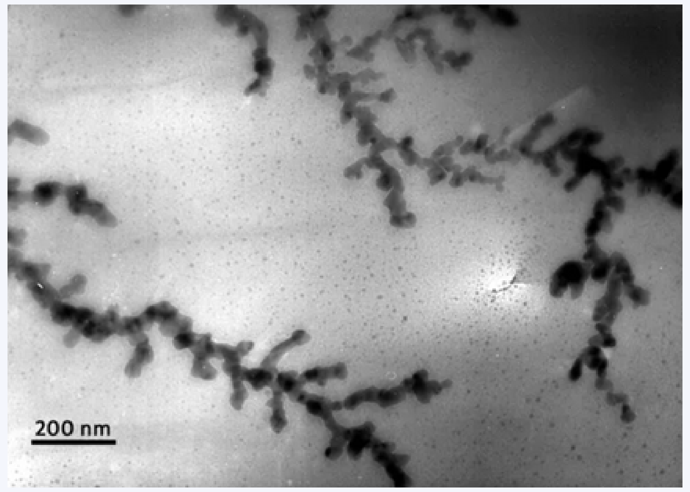
Degradation mechanisms lead to:

- Capacity Fade → Battery stores less energy with time, so devices run shorter. Caused by loss of active lithium and electrode degradation.
- Power Fade → Battery struggles to deliver power quickly. Due to increased internal resistance and electrode damage.
- Safety Risks → Aging raises chances of swelling, overheating, or fire. Triggered by lithium plating, gas formation, or internal shorts.

SEI (SOLID ELECTROLYTE INTERPHASE)

- Formed over the negative Electrode.
- During the very first cycle initial SEI layer is formed ~ 10% initial capacity is lost.
- Acts as a solid electrolyte that blocks stops the further reaction of the electrolyte, however the thickness increases.





Dendrites

Reasons for SEI layer increase:

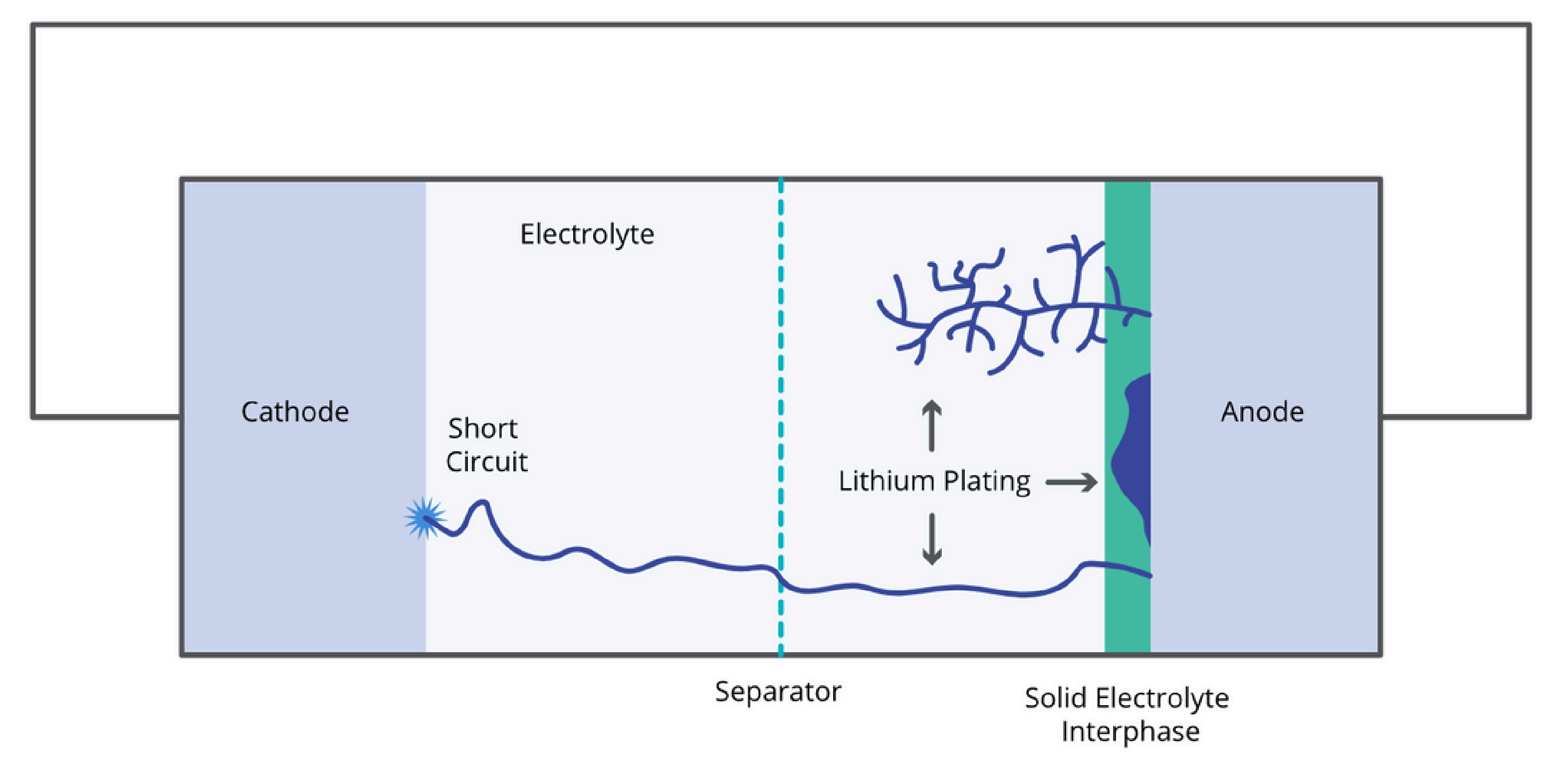
- Solvent diffusion: Electrolyte molecules permeate through pores/defects in existing SEI → react at electrode → form new SEI.
- Electrode cracking: Mechanical stress during cycling exposes fresh graphite surfaces → immediate SEI formation.
- Higher temperature, High currents, High voltage.

Impact of SEI on cell performance:

- Initially, the SEI layer is thin and the performance is not affected much.
- In early stages, the SEI prevents the e^- from anode surface from mixing with the electrolyte.
- As the cell ages, the SEI layer thickens and starts interfering with the major chemical reactions inside the cell.
- Beyond a certain thickness, the anode's ability to absorb the lithium ions reduce. Thus, the ions instead get deposited on the SEI layer, initiating the growth of the dendrites.

LITHIUM PLATING

- Li deposits on the Negative Electrode (NE) surface instead of entering graphite.
- Occurs when Li^+ ions arrive faster than graphite can absorb.
- Can be thermodynamic (NE fully lithiated, no space left) or kinetic (fast charging, cold T).



Reasons for Lithium Plating:

- Low temperature: slows down Li^+ intercalation.
- High SoC: NE nearly full, no room for more Li^+.
- High charging current (fast charging): ions arrive faster than insertion rate.
- High voltage: favors plating over intercalation.
- Local defects: separator flaws, electrode cracks, uneven surface → localized plating.

Impact of Lithium Plating on Cell Performance:

- Reversible plating: some Li can be stripped (recovered).
- Irreversible plating: plated Li reacts with electrolyte → forms SEI or gets isolated as dead lithium.
- Capacity fade: Loss of Lithium Inventory (LLI).
- Power fade: Pore clogging increases resistance.
- Safety risk: Metallic Li can form dendrites, puncturing separator → internal short circuit & fire hazard.

PARTICLE FRACTURE

Reasons behind Particle Fracture:

- It occurs due to large volume changes in electrode materials during lithiation/delithiation, creating stress and cracks

What makes it worse:

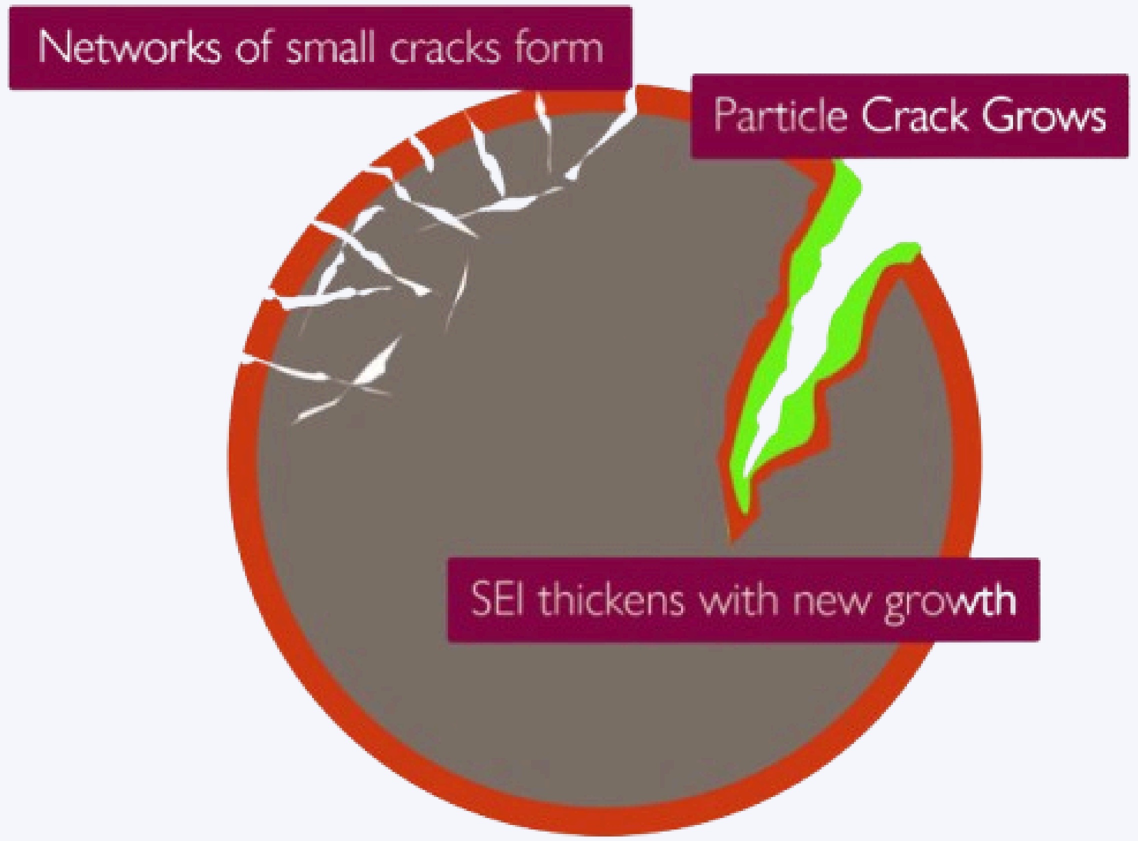
- High/Low temperatures
- very deep charging/discharging
- fast charging
- stresses during making of batteries

What happens after cracking:

- particles lose contact with each other
- newly formed surfaces due to cracking accelerate SEI formation
- Eventually electrode part break off resulting into loss of battery capacity

The Silicon Problem:

- Si which can store much more energy than graphite but it expands almost 4 times when charging
- This again rises the concern about cracking and damage to battery life
- Efforts have made by allowing Si with C to lower this effect but it was still not perfect.



POSITIVE ELECTRODE STRUCTURAL CHANGE AND DECOMPOSITION

- It's a chemistry-dependent degradation, which varies with the chemistry of the positive electrode.

Impact on Battery Degradation:

- Very less active material left results in capacity fade.
- Higher resistances resulting into power fade.
- Gas formation is associated with safety risks.
- Dissolved metals can damage the anode too.

When does this type of decomposition accelerate:

- At high voltage(Higher State of Charge)
- High Temperature
- In the presence of moisture.

What happens inside:

- Its crystal structure changes from a nice layered form to less stable forms(spinel, rock salt, etc.)
- Oxygen atoms can escape, reacting with the electrolyte and forming gases such as CO_2 and CO.
- Some Metals (like Ni,Mn,Co) dissolve into the electrolyte and move to the negative side
- Nickel and Lithium can "swap places" inside the crystal, blocking Li+ movement.
- Acid(from moisture in the electrolyte) makes things worse by dissolving more metals.
- A thin film (pSEI) forms on cathode surface, just like SEI on the anode.

CONCLUSIONS FROM ALL THESE MECHANISMS:

- These mechanisms generally do **not work out individually**; they are complexly **coupled**.
- There are many other mechanisms that play a role in degradation; however, all mechanisms can be grouped into **key categories** based on their impact:
    - Loss of active material
    - Loss of Lithium Inventory
    - Stoichiometric drift due to electrode imbalance
    - Increasing the Internal Resistance
- Degradation broadly depends on:
    - Temperature
    - State of Charge (SoC)
    - Load
    - Materials.
- Experiments with models are essential for lifetime prediction.

DATA DESCRIPTION

- In collaboration with **MIT** and **Stanford**, the **Toyota Research Institute (TRI)** released two high-throughput cycling datasets covering **357** commercial LFP/graphite cells (**A123 Systems APR18650M1A, 1.1 Ah**).
- **Objective:** Study the effect of fast-charging protocols on ageing
- **Charging:**
  - **72 profiles** (single-step) up to ~80% SOC
  - From 80–100%: 1C CC–CV charge at 3.6 V, cutoff = 1/50C–1/20C (≈22–55 mA)
- **Discharge:** Standardized 4C constant current (~4.4 A) to 2.0 V
- **Conditions:** 30 °C chamber, data logged from **cycle 2 to EOL** (80% capacity ≈0.88 Ah)
- **Data recorded:** The dataset contains **in-cycle** measurements (voltage, current, temperature, charge/discharge capacity) and **per-cycle summaries** (capacity, internal resistance, charge time)
- **Cycle life range:** 150–2300 cycles, depending on charging protocol.

Battery Information

Parameters	Value
Nominal Voltage	3.3 V
Nominal Capacity	1.1 Ah
Upper cut-off voltage	3.6 V
Lower cut-off voltage	2.0 V
Cell anode	Graphite
Cell cathode	$LiFePO_4$

ORGANIZATION OF DATA IN THE DATASET

- 124 cells data is divided into 3 batches, each batch contains data of some cells.

1×46 struct with 8 fields

Fields	policy	policy_readable	barcode	channel_id	cycle_life	cycles	summary	Vdlin
1	'3_6C-80PER_3_6C'	'3.6C(80%)-3.6C'	"EL150800460514"	"1"	1190	1×1189 struct	1×1 struct	1000×1 double
2	'3_6C-80PER_3_6C'	'3.6C(80%)-3.6C'	"EL150800460486"	"2"	1179	1×1178 struct	1×1 struct	1000×1 double

- Each cycles column contains the data collected within every cycle of a particular battery.

1×1189 struct with 9 fields

Fields	discharge_dQdV	t	Qc	I	V	T	Qd	Qdlin	Tdlin
1	[]	[]	[]	[]	[]	[]	[]	[]	[]
2	1000×1 double	1087×1 double	1087×1 double	1087×1 double	1087×1 double	1087×1 double	1087×1 double	1000×1 double	1000×1 double
3	1000×1 double	1117×1 double	1117×1 double	1117×1 double	1117×1 double	1117×1 double	1117×1 double	1000×1 double	1000×1 double

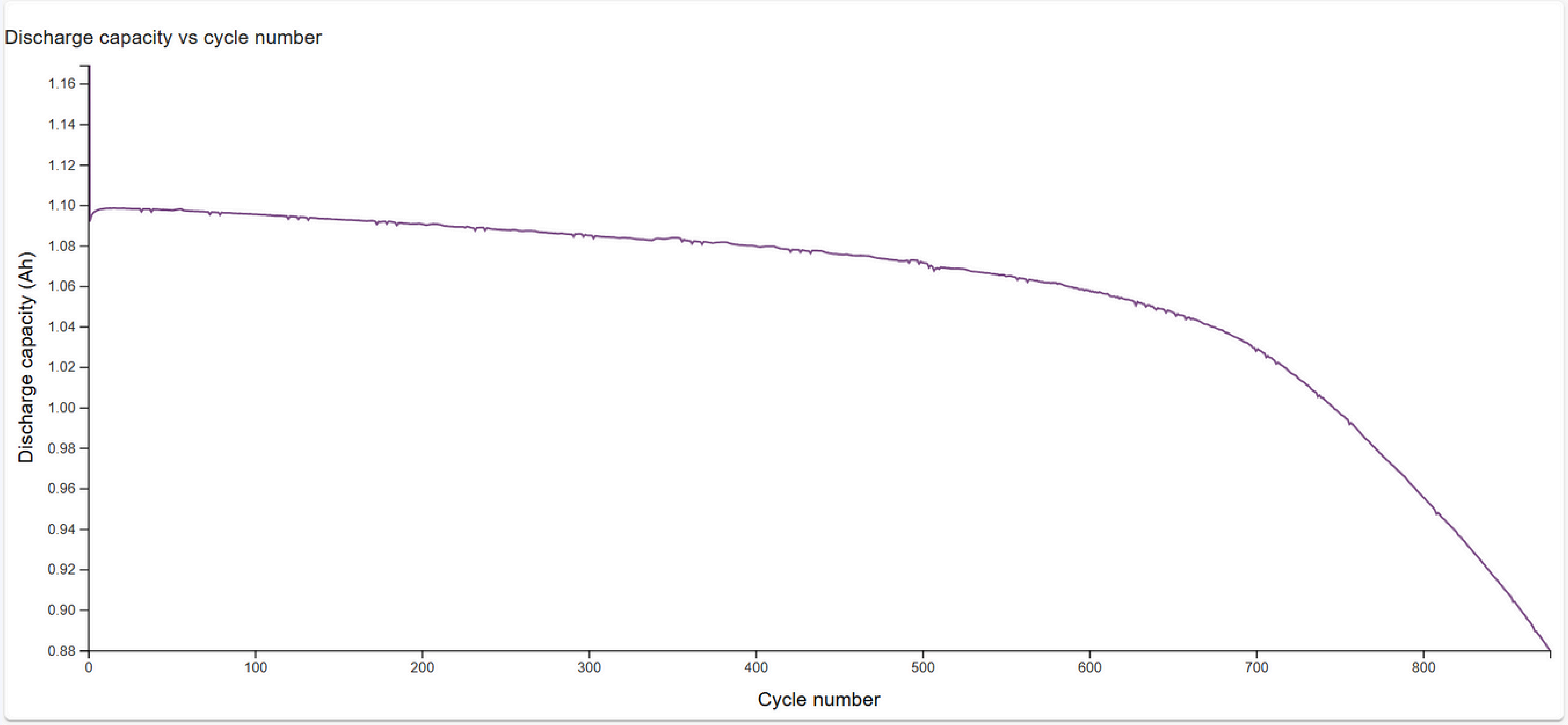
- If we, see the summary column of the data set

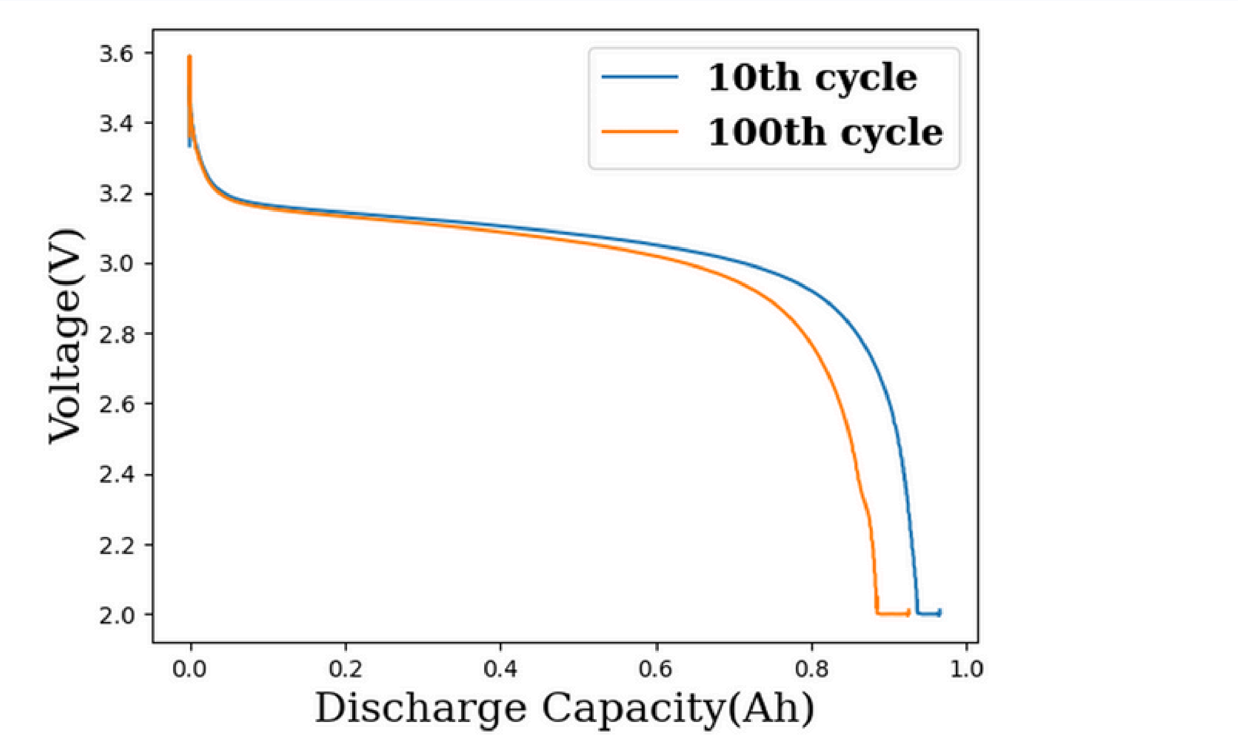
1×1 struct with 8 fields

Field	Value	Size	Class
cycle	1189×1 double	1189×1	double
QDischarge	1189×1 double	1189×1	double
QCharge	1189×1 double	1189×1	double
IR	1189×1 double	1189×1	double
Tmax	1189×1 double	1189×1	double
Tavg	1189×1 double	1189×1	double
Tmin	1189×1 double	1189×1	double
chargetime	1189×1 double	1189×1	double

WHAT THE MODEL IS GOING TO PREDICT FROM THE DATA?

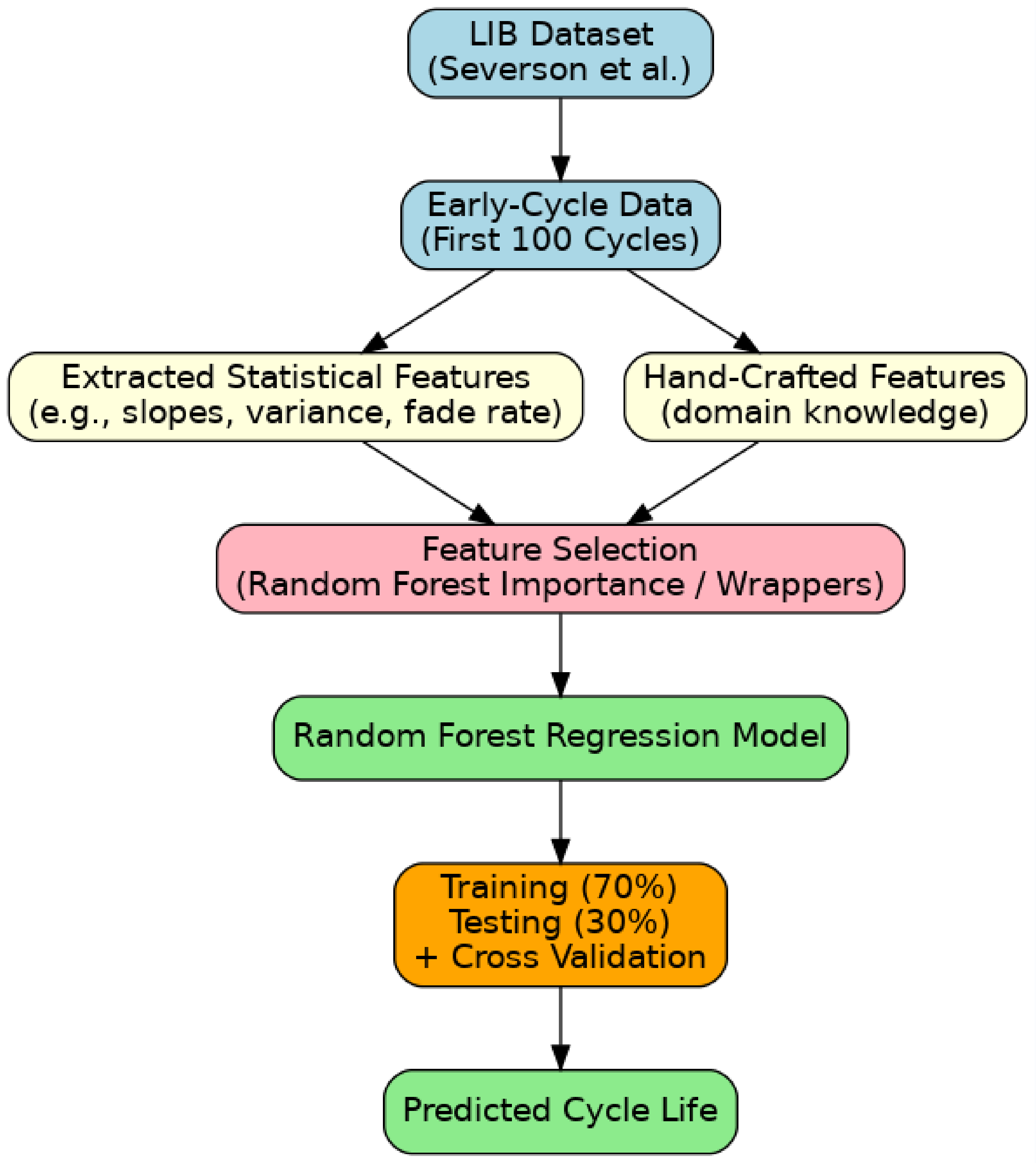
- For each cell, we have the discharge capacity and cycle number data. So, we are going to give only the data of the **first 100 cycles** to the model.
- The model will be provided with the first 100 cycles of discharge capacity data, along with **extracted features,** enabling it to learn patterns and predict the cell’s cycle life in alignment with experimental observations.





WORKING FLOW OF THE MODEL

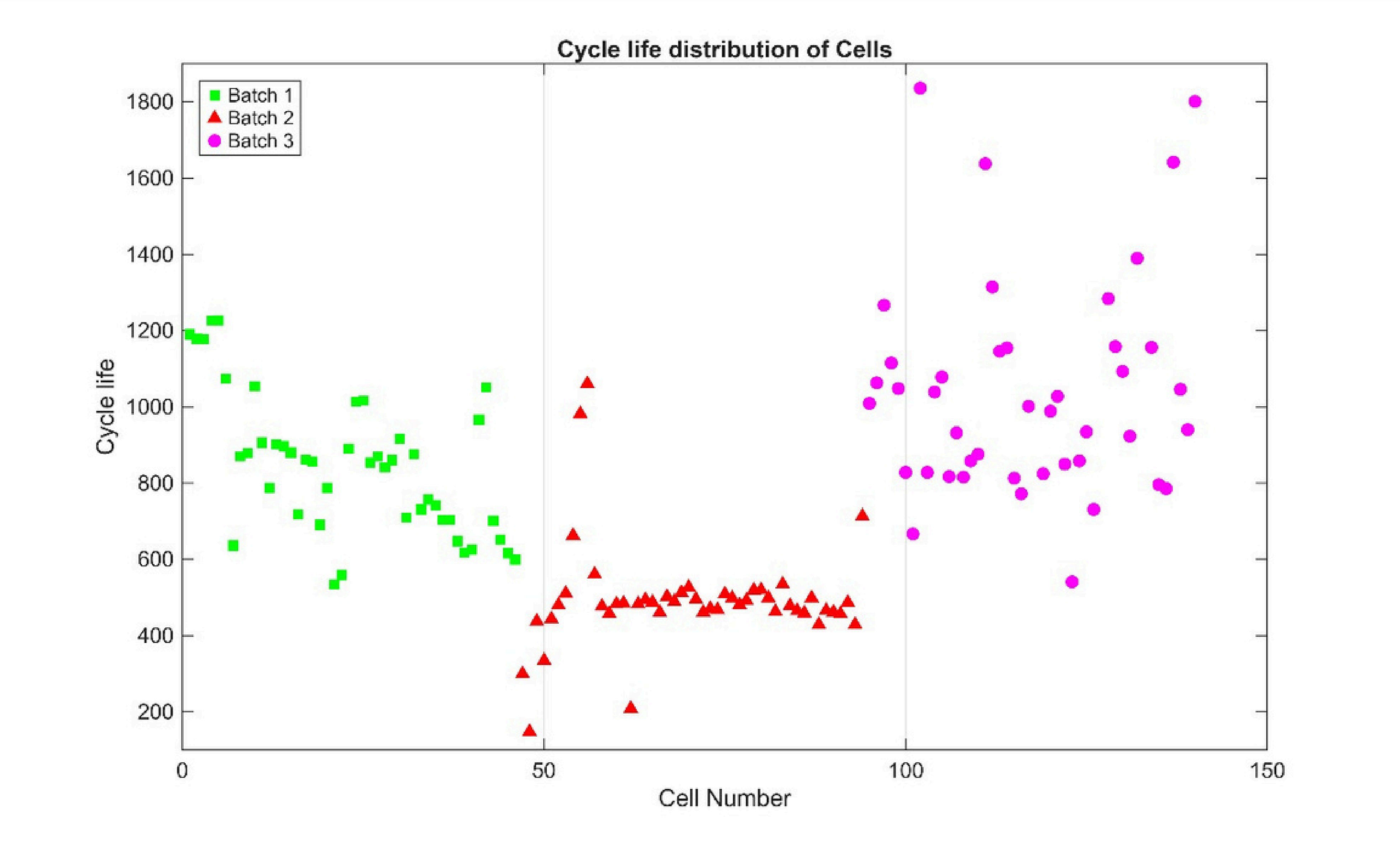
- Below added schematic , is just an overview of how a model works on data.



DATA VISUALIZATION

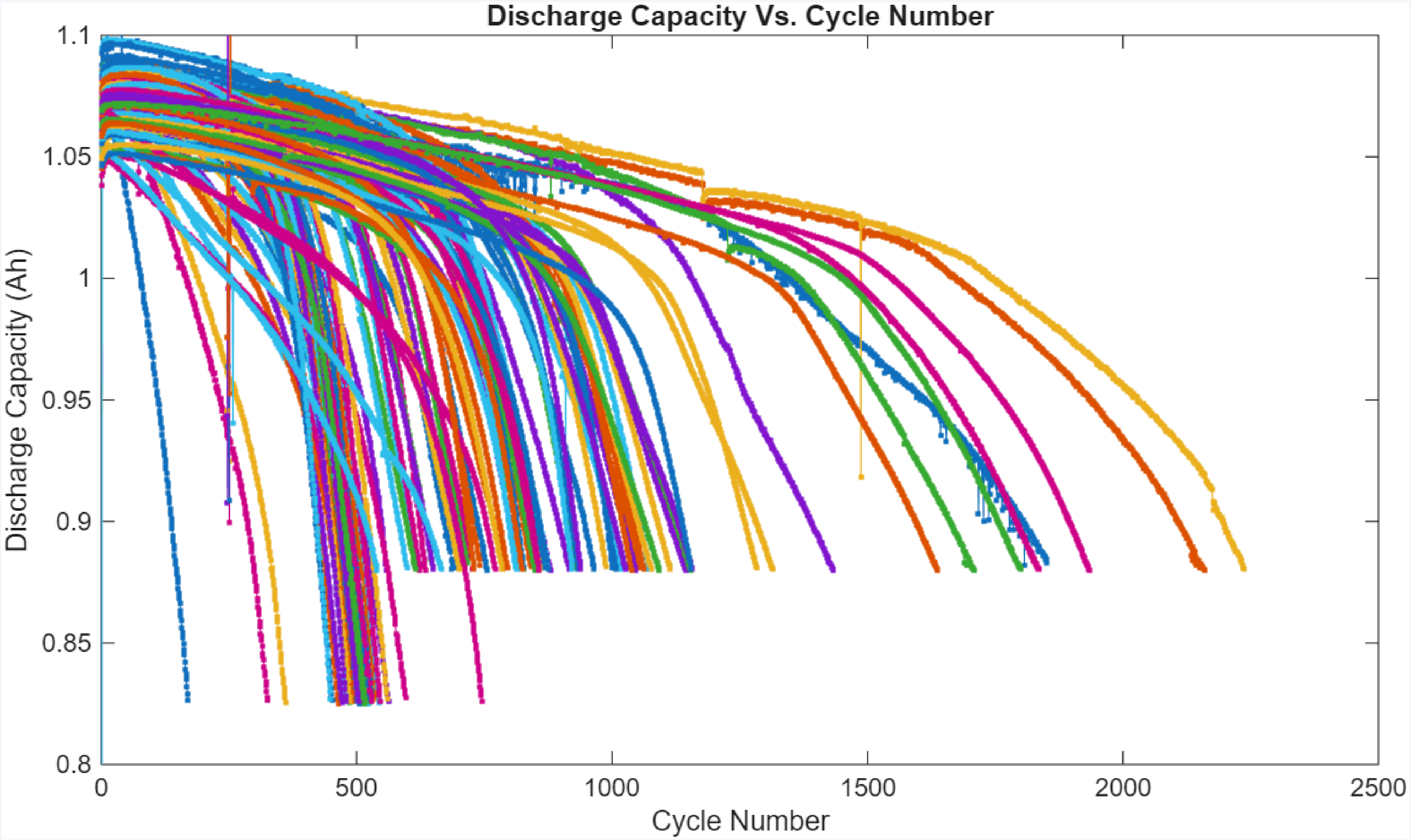
- Data Visualisation gives us better insights into the selected data, hence it is a crucial step before model application

BATCH DATA VISUALIZATION



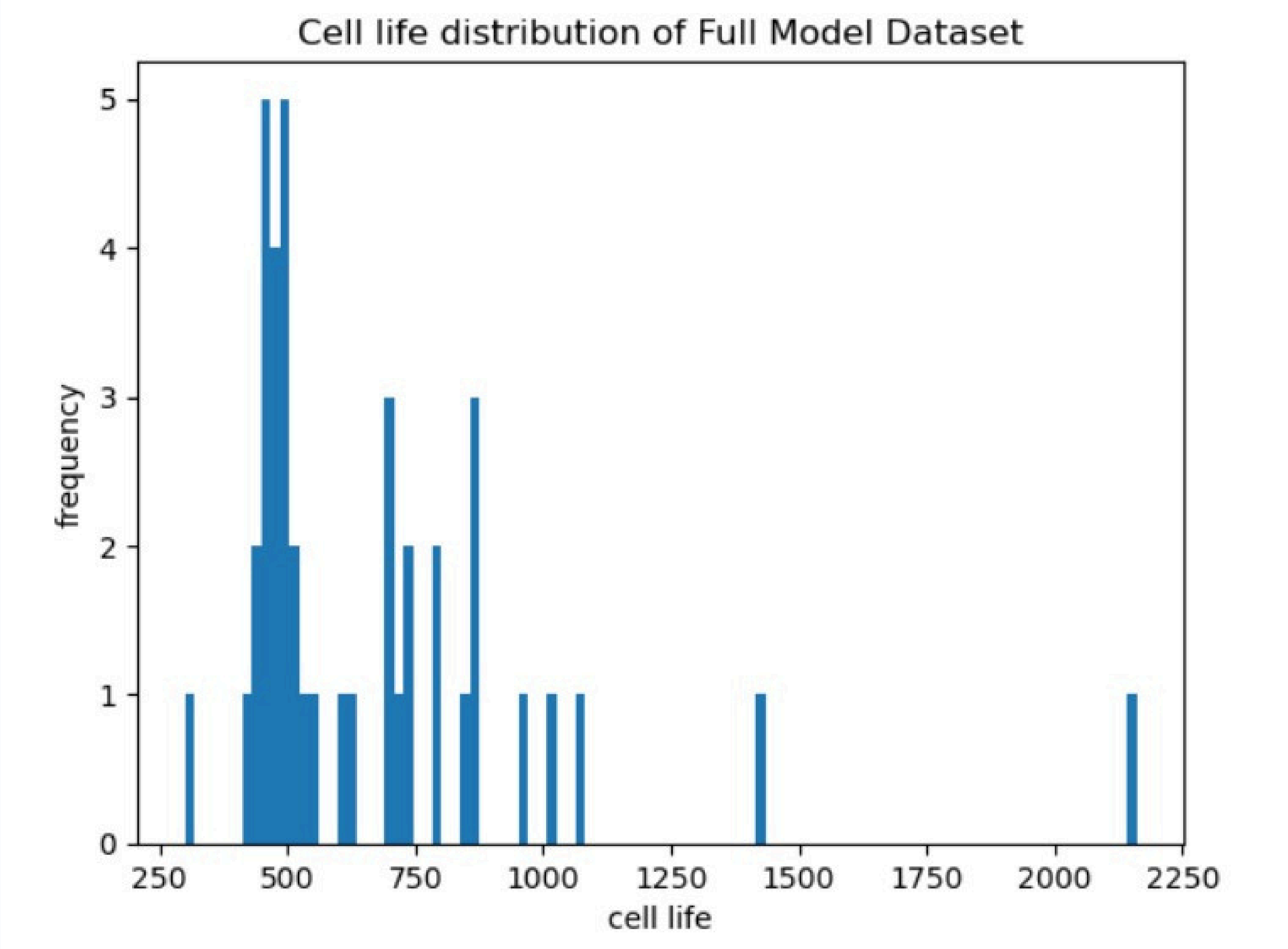
- Total 144 cells (same type) are divided into three batches. The cycle life of each cell in each batch is shown here.

DISCHARGE CAPACITY GRAPH

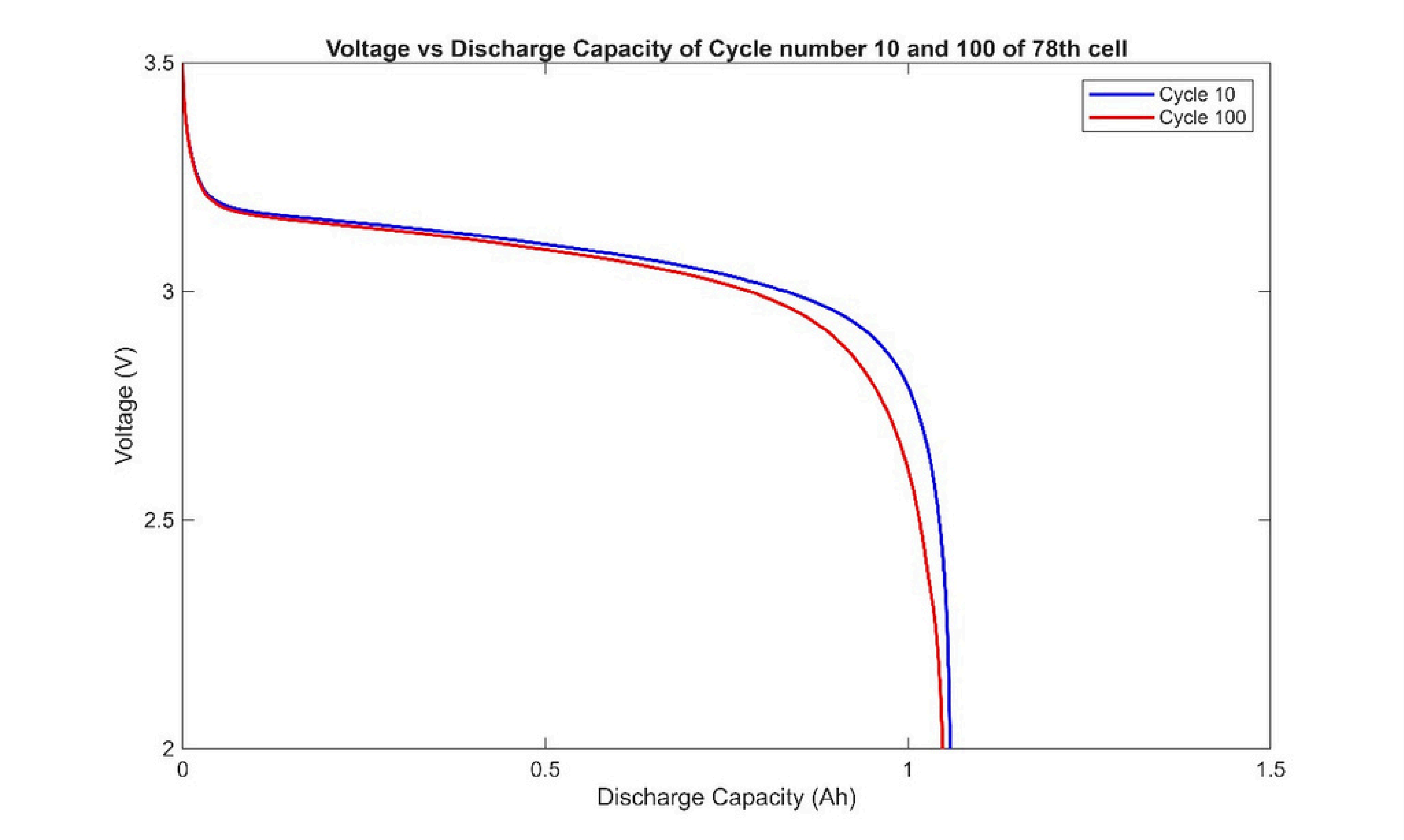


- The most important graph, which tells us the variation of discharge capacity of each cell with number of cycles.
- If all cells are of the same type, then how does the discharge capacity vary in different manner for different cells?
- The reasons behind this have already been discussed in our earlier discussion, that is, the different degradation paths (SEI layer formation, LAM, LLI,etc)
- This gives us an important inference that even small changes in the interior of the battery are caught in this curve very easily, which makes it an important aspect in battery life prediction.

DISTRIBUTION PLOT



VOLTAGE DISCHARGE PLOT



- By comparing curves at early and late cycles, it allows identification of how quickly performance drops, and uncovers hidden degradation mechanisms such as loss of lithium inventory, electrode fracture, or increased internal resistance.
- These features allow scientists and engineers to build models that predict battery aging, recommend charging protocols, or select chemistries that degrade more slowly.

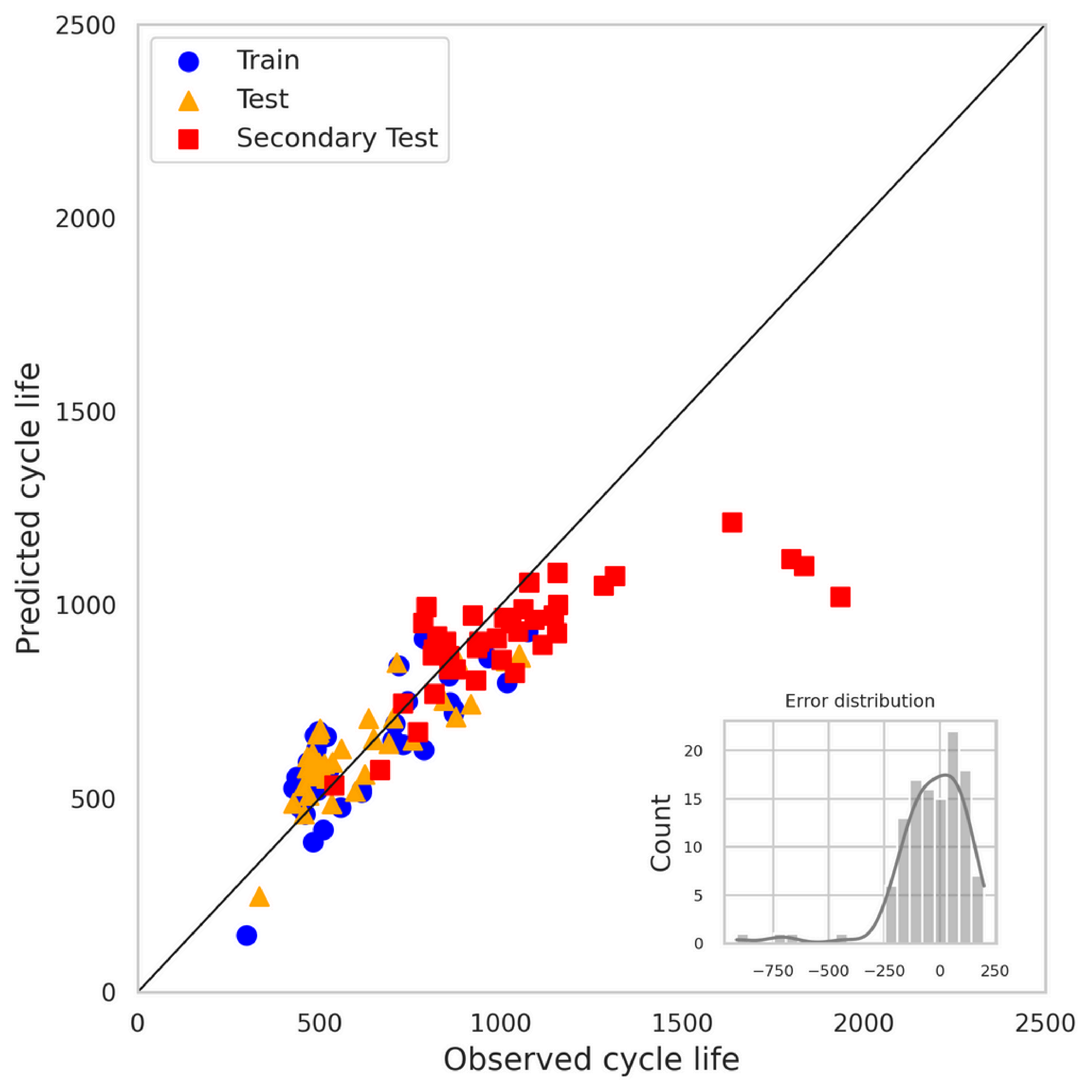
Features Used

	FEATURES		FEATURES
ΔQ100-10(V) features	Minimum	**Other features**	Average charge time, first 5 cycles
	Mean		Maximum temperature, cycles 2 to 100
	Variance		Minimum temperature, cycles 2 to 100
	Skewness		Integral of temperature over time, cycles 2 to 100
	Kurtosis		Internal resistance, cycle 2
	Value at 2V		Minimum internal resistance, cycles 2 to 100
Discharge capacity fade curve features	Slope of the linear fit to the capacity fade curve, cycles 2 to 100		Internal resistance, difference between cycle 100 and cycle 2
	Intercept of the linear fit to capacity fade curve, cycles 2 to 100		
	Slope of the linear fit to the capacity fade curve, cycles 91 to 100		
	Intercept of the linear fit to capacity fade curve, cycles 91 to 100		
	Discharge capacity, cycle 2		
	Difference between max discharge capacity and cycle 2		
	Discharge capacity, cycle 100		

MODEL	NO. OF FEATURES
Variance Model	1
Discharge Model	13
Full Model	20

RESULTS FOR LINEAR REGRESSION

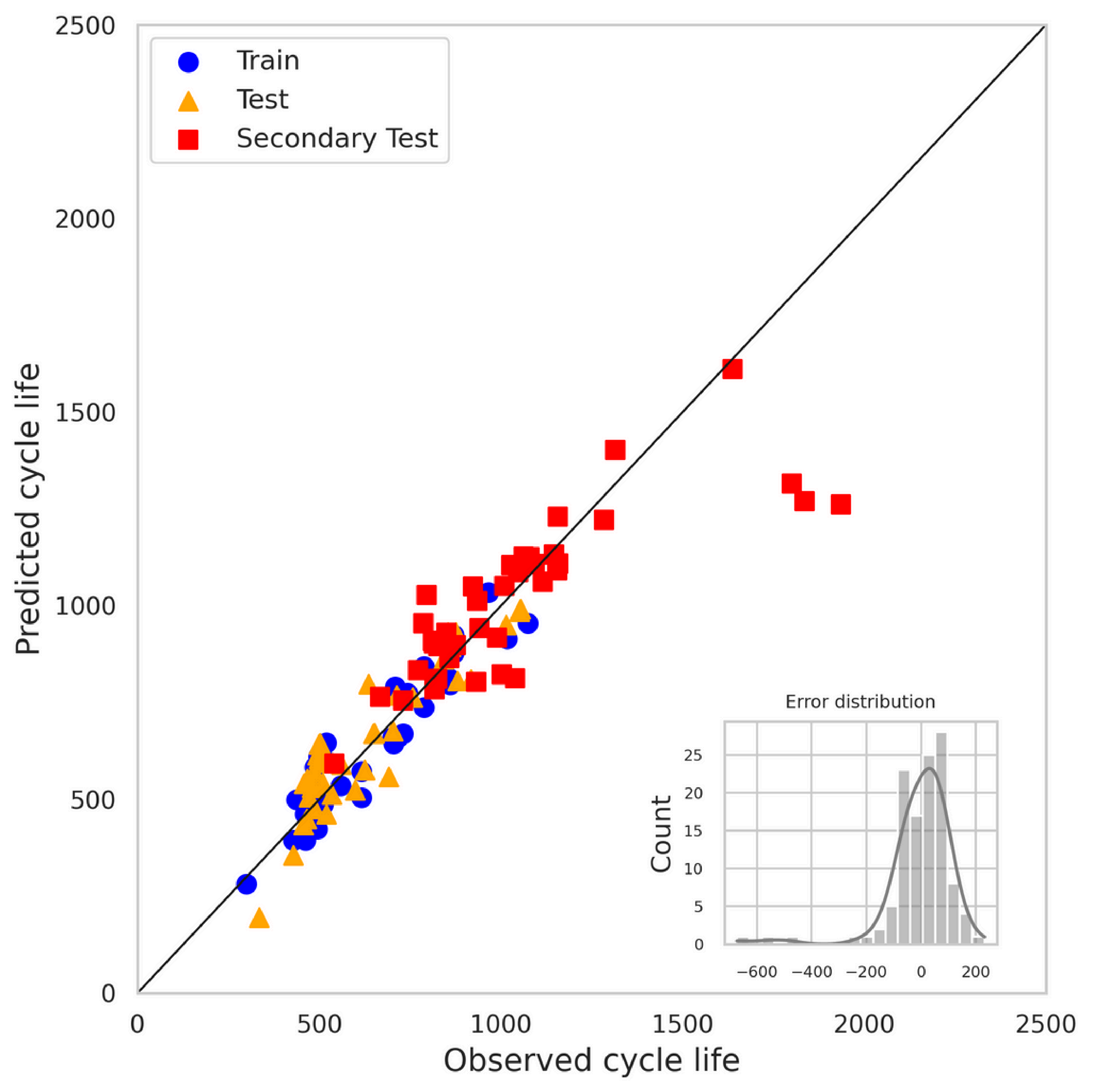
		'Variance' Model
RMSE (cycles)	Train	**108.72**
	Primary test	**104.92**
	Secondary test	**252.49**
Mean percent error (%)	Train	**16.84%**
	Primary test	**14.91%**
	Secondary test	**12.87%**



'VARIANCE' MODEL

RESULTS FOR LINEAR REGRESSION

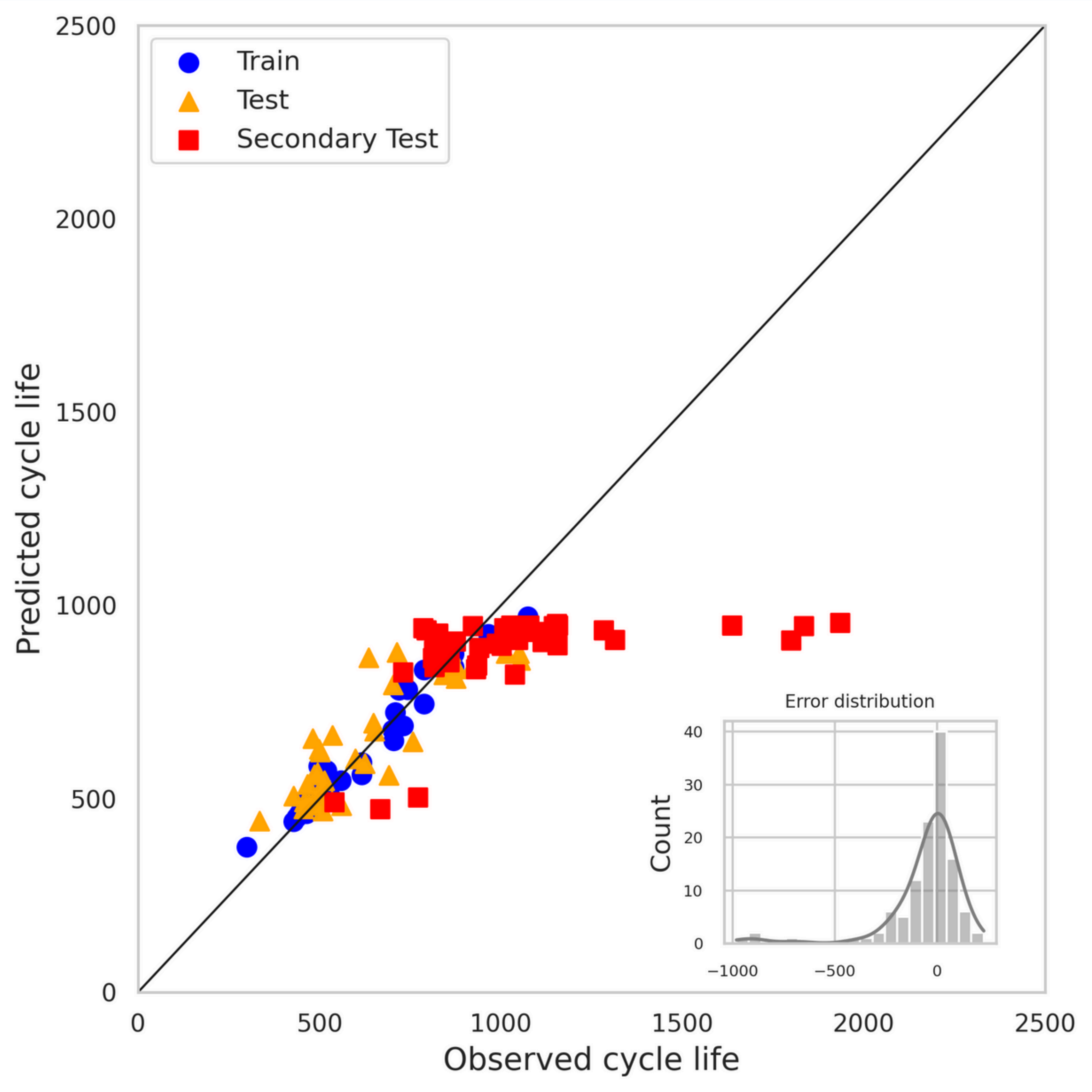
		'Discharge' Model
RMSE (cycles)	Train	**61.06**
	Primary test	**77.42**
	Secondary test	**181.28**
Mean percent error (%)	Train	**8.53%**
	Primary test	**11.66%**
	Secondary test	**9.65%**



'DISCHARGE' MODEL

RESULTS FOR RANDOM FOREST

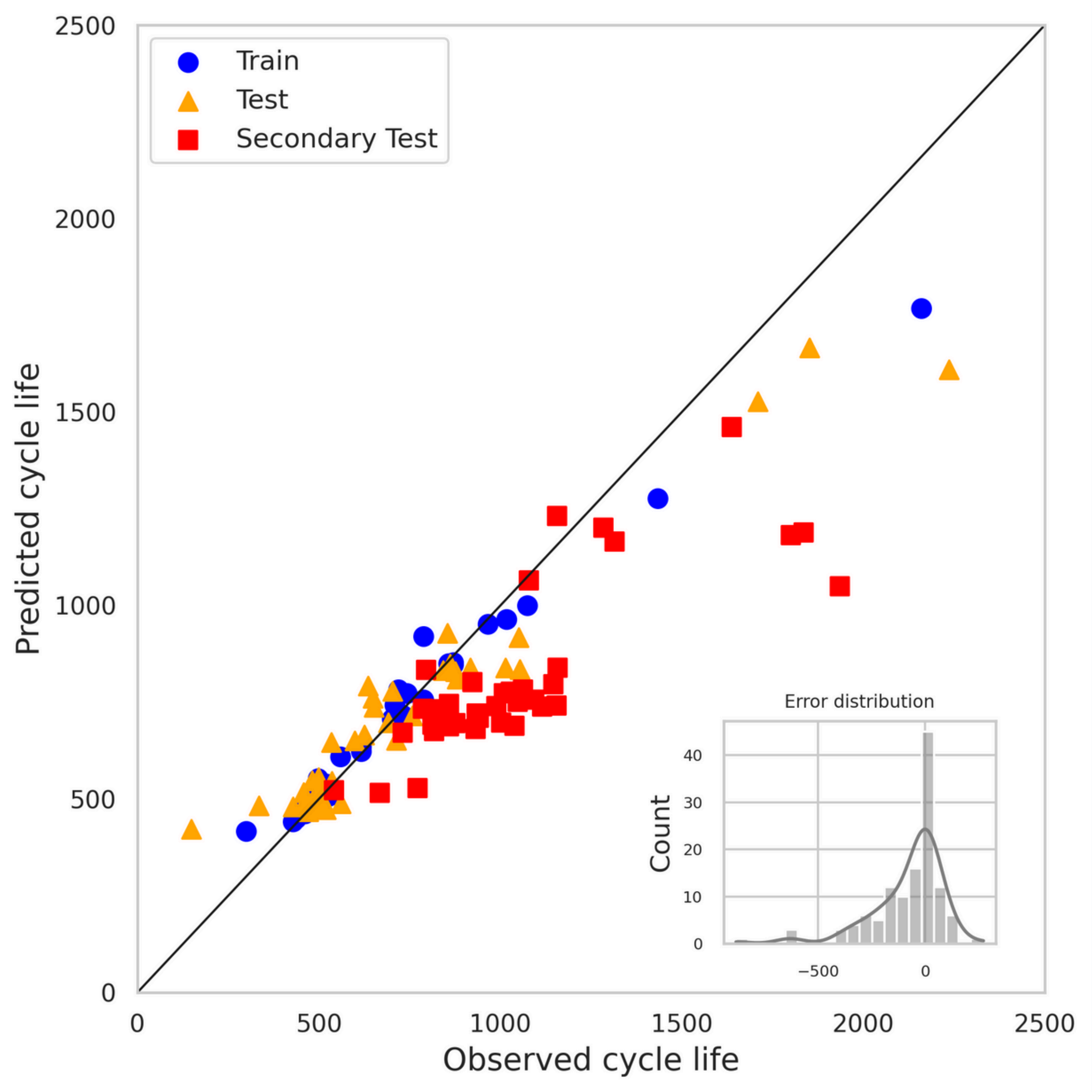
		'Discharge' Model
RMSE (cycles)	Train	39.06
	Primary test	94.64
	Secondary test	312.75
Mean percent error (%)	Train	5.03%
	Primary test	12.09%
	Secondary test	16.49%



'DISCHARGE' MODEL

RESULTS FOR RANDOM FOREST

		'Full' Model
RMSE (cycles)	Train	**75.64**
	Primary test	**133.91**
	Secondary test	**287.68**
Mean percent error (%)	Train	**4.62%**
	Primary test	**13.82%**
	Secondary test	**20.59%**



'FULL' MODEL

FUTURE WORK

- Refinement of both the Linear Regression and Random Forest Model for better predictive performance.
- Understanding the most important features use by the models for prediction.

CODES USED

Link : https://drive.google.com/drive/folders/1_3Ew7p-2iyvLyIHmOUryyLdXYvrfRZQy?usp=drive_link

REFERENCE

- Severson, Kristen A., Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen et al. "Data-driven prediction of battery cycle life before capacity degradation." Nature Energy 4, no. 5 (2019): 383-391.

THANK YOU