# Dataset Details

The synthetic dataset simulates lead-acid battery degradation for TOTO E-rickshaws in Kolkata, comprising 10,000 samples of 7-day sequences for state of health (SOH) prediction, as described in Section IV. Below, we detail its structure, generation, and properties.

## Structure

Each sample represents a 7-day sequence for one E-rickshaw, with 29 features:

- **Client ID**: Unique identifier (1 to 10,000).
- **Features per Day** (Days 1–7):
  - Average Voltage ($V_{\text{avg}}$, V): Daily mean terminal voltage.
  - Current Standard Deviation ($I_{\text{std}}$, A): Variability in current draw.
  - Maximum Temperature ($T_{\text{max}}$, °C): Peak daily temperature.
  - Internal Resistance ($R_{\text{int}}$, $\Omega$): Battery internal resistance.
- **SOH** (%): Target variable, defined as SOH $= \frac{C_{\text{current}}}{C_{\text{nominal}}}$.

The dataset is formatted as a CSV with 29 columns, scalable from the 100-sample subset provided.

## Generation Logic

The dataset is generated based on degradation models (sulfation, corrosion, water loss) under Kolkata's climate (25–42°C), following [1, 2, 3]:

- **Voltage ($V_{\text{avg}}$)**: Starts at 12.7 V, decreases by $\sim$0.1 V/day due to 20% capacity loss over 200 cycles.
- **Current ($I_{\text{std}}$)**: Ranges from 0.3 to 0.9 A, reflecting varied usage (e.g., urban vs. suburban routes).
- **Temperature ($T_{\text{max}}$)**: Peaks at 38–42°C on Day 3, simulating thermal stress events.
- **Internal Resistance ($R_{\text{int}}$)**: Starts at 0.013 $\Omega$, increases by $\sim$0.001 $\Omega$/day due to sulfation (2% per °C above 25°C) and grid deterioration (5–10% annually).
- **SOH**: Ranges from 91.5% to 96.8%, computed based on capacity loss influenced by temperature and usage.

Gaussian noise ($\sigma = 0.05$) is added to ensure realism.

## Non-IID Conditions

The dataset reflects non-IID distributions by varying $I_{\text{std}}$ and $T_{\text{max}}$ across clients, simulating heterogeneous usage patterns (e.g., different routes, driving styles), as requested by Reviewer 1.

## Scalability

The provided 100-sample subset is scalable to 10,000 samples by extending the pattern with additional noise, supporting the federated learning setup (100 clients, 80%–20% train-test split).

## Reproducibility

Generation scripts will be shared at `https://github.com/Surit25/ev/tree/main` in a future version, ensuring transparency per Reviewer 3's feedback.

## Usage

The dataset supports TALNet training, achieving a MAPE of 0.8% (Section V). It aligns with the pilot study results (MAPE of 0.82%, 20% lifespan extension) and Figure 4 (SOH trends).

# References


[1] Y. Jiang and Z. Song, "Sulfation Effects in Lead-Acid Batteries," *Electrochimica Acta*, vol. 412, 2022.

[2] L. Chen et al., "Grid Deterioration in Lead-Acid Batteries," *Journal of Power Sources*, vol. 589, 2024.

[3] D. Kumar et al., "Impact of High Temperatures on Lead-Acid Battery Degradation," *Journal of Energy Storage*, vol. 55, 2023.