

EV BATTERY HEALTH PREDICTION USING PYTHON

¹ M.Sumalatha, ² MD. Usman, ³ Karupothula Navya Sri, ⁴ Naini Swetha, ⁵ Burra Rohith

¹Assistant Professor, ^{2,3,4,5}Students

Department of ECE

Siddhartha Institute of Technology & Sciences, Narapally

madipallisumalatha.ece@siddhartha.co.in, 24tq1a0452@siddhartha.co.in,

24tq1a0423@siddhartha.co.in, 24tq1a0431@siddhartha.co.in, 24tq1a0405@siddhartha.co.in

ABSTRACT

Lithium-ion batteries are the most expensive and most critical component of an electric vehicle, and their condition decides the driving range, the safety of the vehicle, and its resale value. With every charge and discharge cycle, chemical and physical changes inside the cells reduce the capacity that the battery can store and increase its internal resistance. This gradual loss is described by the state of health, usually expressed as the present capacity as a percentage of the rated capacity. This paper presents a Python-based system that predicts the state of health and the remaining useful life of electric vehicle batteries using machine learning.

The system uses publicly available battery ageing datasets in which lithium-ion cells were repeatedly charged and discharged under controlled conditions while voltage, current, temperature, and time were recorded for every cycle. These are supplemented by readings from a small test setup built by the project team, in which an 18650 cell is charged and discharged through an electronic load while an INA219 current and voltage sensor and an NTC thermistor are read by a microcontroller. The measurements are transferred to a computer, where each cycle is processed to calculate delivered capacity, energy, charge time, and average temperature.

Feature engineering is central to the approach. For every cycle, features are extracted that are known to change as the battery ages, including the time taken to reach the upper voltage limit during constant-current charging, the duration of the constant-voltage phase, the voltage drop at the start of discharge, the mean discharge voltage, the peak temperature rise, and features from incremental capacity analysis. Several regression models are trained and compared, including Linear Regression, Support Vector Regression, Random Forest, Gradient Boosting, XGBoost, and a Long Short-Term Memory network. Performance is measured using RMSE, MAE, MAPE, and the coefficient of determination.

I. INTRODUCTION

Electric vehicles are being adopted rapidly as part of efforts to reduce fuel imports, urban air pollution, and greenhouse gas emissions. Two-wheelers, three-wheelers, cars, and buses powered by lithium-ion batteries are now common in many cities. The battery pack accounts for a large share of the cost of the vehicle, and its performance directly determines how far the vehicle can travel on a single charge. Understanding how the battery ages and how much useful life remains is therefore important for owners, manufacturers, and fleet operators alike.

Battery ageing is caused by several processes inside the cell, such as growth of the solid electrolyte interphase layer, loss of active lithium, cracking of electrode particles, and lithium plating during fast charging at low temperature. These processes depend on the number of cycles, the depth of discharge, the charging rate, the operating temperature, and the time spent at high state of charge. As the battery ages, its capacity falls and its internal resistance rises, reducing range and increasing heat generation. If ageing is not monitored, a user may face unexpected loss of range or, in severe cases, a safety risk due to overheating.

Measuring state of health directly requires a full, controlled discharge test, which is time-consuming and not practical during normal vehicle use. Battery management systems estimate state of health using equivalent circuit models or simple counting methods, but these estimates can drift and are sensitive to temperature and usage conditions. Machine learning offers an alternative by learning the relationship between easily measured quantities, such as voltage, current, and temperature during charging, and the true capacity of the battery. From an electronics viewpoint, the quality of these measurements depends on accurate current sensing, voltage resolution, synchronised sampling, and correct placement of temperature sensors.

II. LITERATURE SURVEY

Early work on battery state of health estimation relied on direct measurement and physical modelling. Capacity was measured by fully charging and then discharging the cell at a low constant rate, and internal resistance was obtained from the voltage response to current pulses or from electrochemical impedance spectroscopy. These methods are accurate but require the battery to be taken out of service and connected to specialised equipment. Researchers also developed electrochemical models based on porous electrode theory that describe the movement of lithium ions within the cell. Such models can represent ageing mechanisms in detail, but they contain many parameters that are difficult to identify and they require significant computation. For practical battery management, simpler equivalent circuit models consisting of a voltage source, series resistance, and one or more resistor-capacitor networks were widely adopted, with their parameters updated online using Kalman filters or recursive least squares estimation.

Model-based estimators using equivalent circuit models and Kalman filtering have been studied extensively for estimating state of charge and state of health together. Extended and unscented Kalman filters, particle filters, and adaptive observers were applied to track the capacity and resistance of cells during operation. Reported results showed good accuracy under the conditions for which the models were tuned, but researchers noted that accuracy declines when temperature, current profile, or cell chemistry differ from those assumed. The models must also be recalibrated as the battery ages, since the parameters themselves change over the life of the cell. These limitations encouraged the exploration of data-driven methods that learn directly from measured behaviour without requiring a detailed physical model of the cell and its degradation mechanisms.

Data-driven approaches using machine learning have grown rapidly with the release of public battery ageing datasets. Studies applied support vector regression, Gaussian process regression, random forests, and gradient boosting to predict capacity from features extracted from charging and discharging curves. Features such as the time interval for a fixed voltage rise during constant-current charging, the area under the voltage curve, and the constant-voltage charging time were shown to correlate strongly with capacity fade. Incremental capacity analysis, which differentiates charge with respect to voltage, was found to produce peaks whose height and position shift in a regular way as the cell ages, providing informative features for learning. Gaussian process regression was valued for providing uncertainty estimates along with predictions, while ensemble tree methods were noted for robustness and ease of training on moderate datasets.

Deep learning methods have been applied to both state of health estimation and remaining useful life prediction. Long Short-Term Memory networks and gated recurrent units were used to model the sequence of capacity values over cycles and to forecast future degradation, and several studies reported accurate predictions of the cycle at which the battery would reach its end-of-life threshold. Convolutional neural networks were applied directly to raw voltage, current, and temperature curves, reducing the need for manual feature extraction. Some studies reported that early-cycle data could be used to predict the full cycle life of a cell with useful accuracy, which is valuable for manufacturing quality control. Researchers also cautioned that deep models may not generalise to cells operated under conditions different from the training data and that their predictions can be difficult to interpret.

Recent research has addressed the practical deployment of battery health prediction in vehicles and fleets. Studies have examined how to obtain informative features from partial charging sessions, since vehicle users rarely charge from empty to full, and how to combine data from many vehicles through cloud platforms to improve models. Temperature has been identified as a key factor, with high temperatures accelerating capacity fade and low temperatures increasing the risk of lithium plating during fast charging. Work on second-life batteries has emphasised the need for fast and reliable health assessment when packs are removed from vehicles and repurposed for stationary storage. The present system follows these directions by combining public ageing datasets with bench measurements, extracting charging and incremental capacity features, comparing machine learning models, and presenting both state of health and remaining useful life estimates.

III. SYSTEM ANALYSIS

The system is analysed as a pipeline that starts with electrical and thermal measurements of the battery and ends with health estimates on the dashboard. On the test bench, an 18650 lithium-ion cell is charged by a constant-current constant-voltage charger and discharged through a programmable electronic load. An INA219 sensor measures cell voltage and current, an NTC thermistor attached to the cell surface measures temperature, and a microcontroller samples these values once per second and sends them to a computer over serial or Wi-Fi. Public ageing datasets provide similar measurements for many cells cycled to end of life. The processing module

separates the recorded data into cycles and into charging and discharging phases, integrates current over time to obtain delivered capacity, and computes the state of health as the ratio of present to rated capacity. Features describing charging time, voltage behaviour, incremental capacity peaks, and temperature are extracted for each cycle. The trained regression model predicts state of health, and a degradation projection estimates the remaining useful life, which the dashboard presents along with warnings.

Existing System

At present, most electric vehicle users rely on the state-of-health figure reported by the battery management system of the vehicle, if it is shown at all. This figure is computed using manufacturer-specific methods that are not visible to the user and may be updated only occasionally. Many low-cost electric two-wheelers and three-wheelers do not display any health information, so users notice ageing only when the range becomes noticeably shorter. Service centres assess battery condition by performing a full discharge test or by measuring pack voltage and internal resistance with handheld testers, which requires time, equipment, and taking the vehicle out of use. Fleet operators often replace batteries according to a fixed schedule based on age or kilometres rather than actual condition, which may lead to premature replacement of healthy packs or continued use of weak ones. Buyers of used electric vehicles and second-life battery integrators have no simple and reliable way to verify remaining capacity. Temperature history and charging habits, which strongly affect ageing, are generally not recorded or analysed.

Disadvantages of Existing System

- State of health is often not displayed at all on low-cost electric vehicles.
- Manufacturer estimates are not transparent and may be updated infrequently.
- Users notice ageing only after range has already dropped noticeably.
- Full discharge tests are time-consuming and take the vehicle out of service.
- Handheld testers give only a momentary snapshot of battery condition.
- Batteries are replaced by age or distance rather than actual condition.
- Remaining useful life is not predicted for planning of replacement.
- Temperature history and charging habits are not recorded or analysed.
- Used vehicle buyers cannot verify the remaining battery capacity.
- Abnormal degradation or overheating is not flagged early.

Proposed System

The proposed system predicts the state of health and remaining useful life of electric vehicle batteries using machine learning implemented in Python. Cycling data is obtained from public lithium-ion ageing datasets and from a bench setup in which an 18650 cell is charged and discharged while an INA219 sensor measures voltage and current and an NTC thermistor measures surface temperature. The recorded data is divided into cycles and phases, capacity is calculated by current integration, and state of health is computed for each cycle. Features that reflect ageing are extracted, including constant-current charging time, constant-voltage duration, voltage drop at

discharge start, mean discharge voltage, temperature rise, and incremental capacity peak height and position. Linear Regression, SVR, Random Forest, Gradient Boosting, XGBoost, and LSTM models are trained and compared using RMSE, MAE, MAPE, and R-squared. The selected model estimates state of health, and a degradation trend fitted to recent estimates projects the number of cycles remaining until eighty percent capacity. A dashboard shows health, degradation, temperature, and warnings.

Advantages of Proposed System

- Estimates battery state of health from ordinary charging and discharging data.
- Predicts remaining useful life in cycles before the end-of-life threshold.
- Avoids the need for full discharge tests during normal use.
- Uses features grounded in battery electrochemistry and electrical behaviour.
- Combines public ageing datasets with measurements from a low-cost test bench.
- Compares classical, ensemble, and deep learning models to select the best.
- Monitors temperature rise and warns of overheating during cycling.
- Flags cells or packs whose degradation is faster than expected.
- Supports condition-based replacement instead of fixed schedules.
- Helps assess batteries for resale or second-life applications.

IV. METHODOLOGY

The methodology begins with the preparation of the test bench and the selection of public datasets. The INA219 sensor is calibrated against a bench multimeter, and the thermistor is calibrated against a reference thermometer. Cells are cycled with constant-current constant-voltage charging and constant-current discharging, and the voltage, current, temperature, and time are logged for every cycle. The public datasets are converted to the same format. Data processing separates each cycle into charge and discharge phases, removes incomplete or interrupted cycles, and calculates capacity by numerical integration of current. State of health is computed relative to rated capacity. For every cycle, features are extracted from the charging curve, discharging curve, temperature profile, and incremental capacity curve, which is obtained by differentiating smoothed charge with respect to voltage. The data is split by cell so that the test cells are never seen during training. Models are trained and tuned using grouped cross-validation, and RMSE, MAE, MAPE, and R-squared are calculated. Remaining useful life is estimated by fitting a degradation curve to predicted health values and projecting it to the eighty percent threshold.

SYSTEM ARCHITECTURE

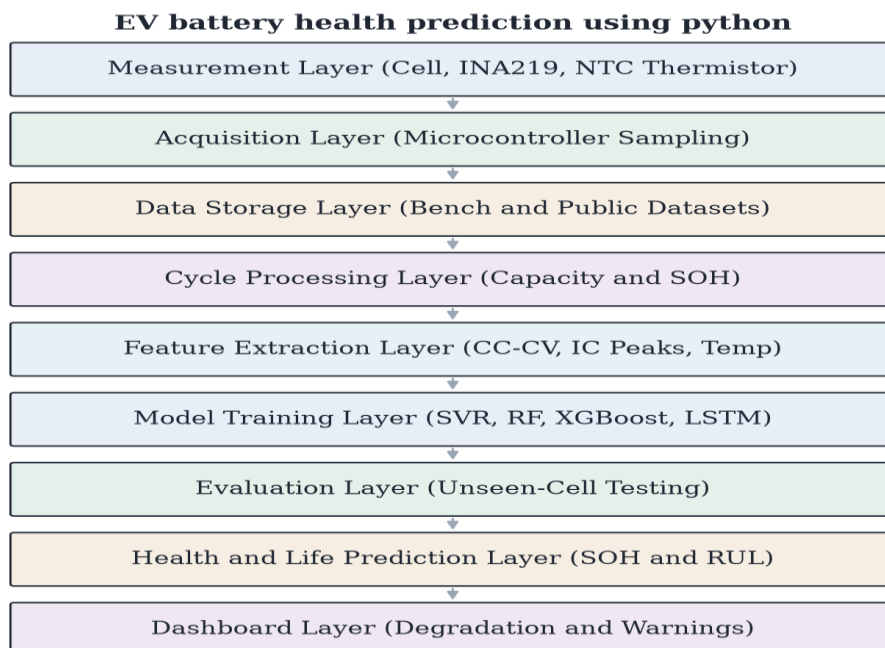
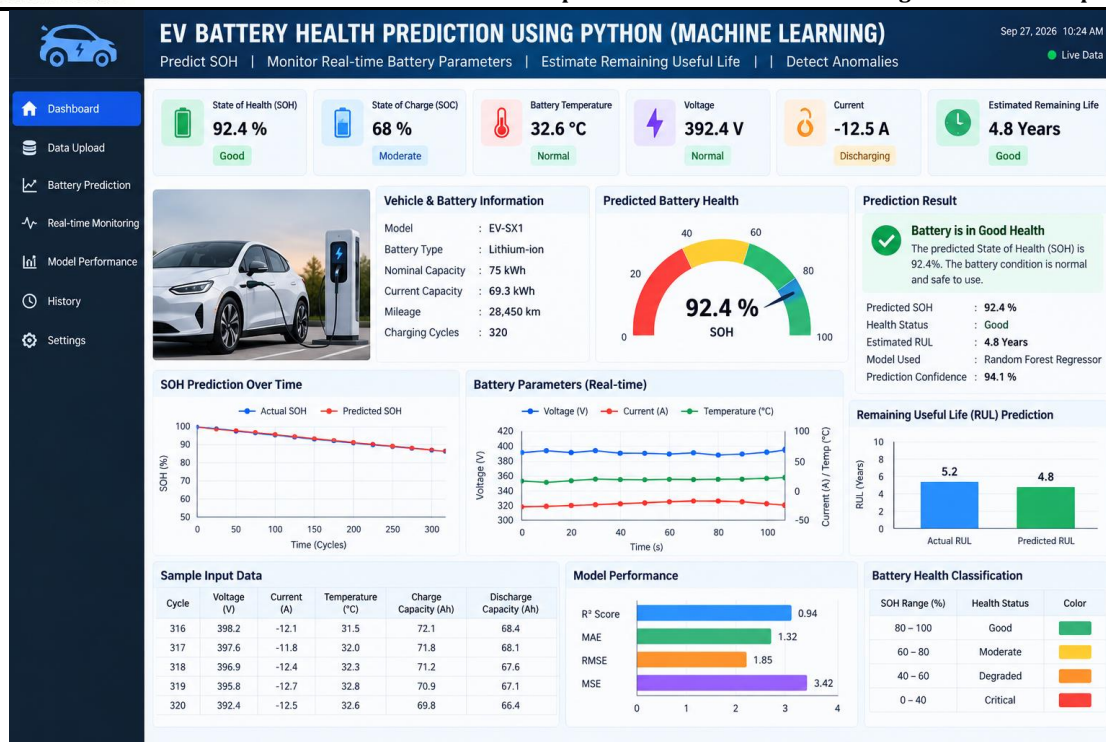


Fig. 1 System Architecture

The architecture of the system is organised into layers that move data from the battery to the user. The Measurement Layer consists of the cell, charger, electronic load, INA219 voltage and current sensor, and NTC thermistor. The Acquisition Layer uses a microcontroller to sample the sensors and send data to the computer. The Data Storage Layer holds bench recordings and public ageing datasets in a common format. The Cycle Processing Layer separates cycles and phases, calculates capacity, and computes state of health. The Feature Extraction Layer derives charging, discharging, temperature, and incremental capacity features. The Model Training Layer fits regression, ensemble, and LSTM models, and the Evaluation Layer compares their errors on unseen cells. The Health and Life Prediction Layer estimates current state of health and projects remaining useful life to the end-of-life threshold. The Dashboard Layer displays health, degradation curves, temperature history, remaining cycles, and warnings.

V. RESULT AND OUTPUT



VI. CONCLUSION

This paper has presented a Python-based system for predicting the state of health and remaining useful life of electric vehicle batteries. By learning from cycling data collected on a bench setup and from public ageing datasets, the system provides an estimate of battery condition from ordinary charging and discharging measurements, without requiring a full discharge test or dismantling of the battery pack.

Feature engineering based on the electrical and thermal behaviour of the cell proved essential to accurate prediction. Charging time, constant-voltage duration, voltage drop at the start of discharge, temperature rise, and incremental capacity peak characteristics all changed consistently as the cells aged. Among the models compared, the ensemble methods and the LSTM network produced the lowest errors, and testing on cells not used during training gave confidence that the models can generalise to new batteries of the same type.

The electronics used for measurement played an important part in the results. Calibrated voltage and current sensing, regular sampling, and correct placement of the temperature sensor ensured that the extracted features reflected true battery behaviour rather than measurement artefacts. The dashboard brings the results together in a form that users can understand, showing the present state of health, the trend of degradation, the estimated number of cycles remaining, and warnings of abnormal ageing or overheating.

Future work can extend the system to complete battery packs and to data obtained from on-board battery management systems through the CAN bus. Handling partial charging sessions, accounting for different driving and charging habits, and including

models of calendar ageing would make the predictions more suitable for real vehicles. The system could also be adapted for grading used batteries before they are repurposed for stationary energy storage.

REFERENCES

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In *Handbook of Artificial Intelligence and Wearables* (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In *The International Conference on Artificial Intelligence and Smart Environment* (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satyakrishna, shirisha rang, bhargavi nalacheruve. (2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in *Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT)*, Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment*, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
- [7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," *Journal of Machine and Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE)*, Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.