Integrated Machine Learning–Based SOC Prediction, Regenerative Energy Optimization, and Battery Health Assessment for Urban Logistics Electric Vehicles

Mavuri Arunakumari
Department of Electrical & Electronics Engineering ,
Vignan's Institute of Information Technology (A), Besides VSEZ, Vadlapudi, Duvvada, Visakhapatnam, Andhra Pradesh-530049, India
arunamavuri@vignaniit.edu.in

Parvathi
Department of Electrical & Electronics Engineering
Ballari institute of technology and management, Ballari
bhumiparvathi830@gmail.com

Ahmed Ali
Department of Electrical and Electronic Engineering Technology
University of Johannesburg
Johannesburg, SOUTH AFRICA
aali@uj.ac.za

Ramakrishna S S Nuvvula
NITTE (Deemed to be University)
NMAM Institute of Technology
Karnataka, Department of Electrical and Electronics Engineering, Department of Electrical and Electronics Engineering

***Abstract*—Electric vehicles for urban logistics are still facing major problems regarding energy management and maximizing the battery life. This paper, in fact, discloses an entire machine learning (ML) oriented framework, including a State of Charge (SOC) predictor and an energy management system for light commercial L5, class vehicles. ESOGU, ML5EV which is the origin of the dataset applied here, holds a total of 562 data points, each one corresponding to one second of record, for a 2.0 km urban delivery route under 300 kg load. These points have been taken at one, second intervals. A total of six different models were chosen for the experiment, and their performances were compared. The models targeted were Linear Regression, XGBoost, Random Forest, LSTM, GRU, and Stacked Ensemble. Among them, Linear Regression excelled with a remarkably low RMSE of 0.003%, which is 667 times better than the target, an R of 0.879, and an inference time of 0.007 ms making it deployable in real, time applications." On the other hand, the analysis of the energy consumption breakdown indicated that only 10.1% of the moments on the road result in 42.3% of the total energy consumption, whereas the energy recovered through regenerative braking was 13.3% of the energy used and its maximum potential could.**

Keywords: Electric vehicles; Urban logistics; State of Charge estimation; Machine learning; Energy management system.

I. INTRODUCTION

1.1 Motivation and Background

Urban freight transport plays a vital role in the logistics infrastructure of modern cities. Nevertheless, it is one of the main factors to blame for greenhouse gas emissions and urban air quality deterioration. In other words, the logistics sector is under pressure to decrease its carbon footprint as cities worldwide are experiencing an increase in delivery demands caused by a boom in e, commerce and widespread adoption of just, in, time supply chain models. Among various proposals to decarbonize last, mile delivery, electric vehicles (EVs) have been considered as a proper solution from a technology standpoint, a green solution that produces no emissions at the point of use, and still, offers a level of operational flexibility suitable for dense urban areas.

The transition to electric urban logistics fleets is accompanied by a variety of challenges that concern not only vehicle electrification but also, for instance, battery management, particularly when one looks at the adoption of L5, class cargo vehicles for commercial freight operations. Battery management is the only lever to keeping the operation side of things feasible, given the fact that the one factor which most drastically influences vehicle range estimation, scheduling of charging, and finally, fleet operational reliability, is the precision of State of Charge (SOC) prediction. Therefore, it is the energy management systems that are regarded as the essential elements that allow one to get the most out of vehicle utilization while simultaneously keeping the total cost of ownership at the minimum level, which is by far the most crucial aspect for a fleet operator when deciding to commit to a new technology. Indeed, urbanization of logistics operations increases the severity of these problems twofold because of the nature of operation involving frequent acceleration and deceleration cycles, variable payload configurations, complex route topographies, and tight delivery time windows.

Battery capacity degradation caused by the aging of the cells and the stress experienced during the urban delivery cycles means one of the major parts of an electric vehicle's total cost of ownership since mounting of the lithium, ion battery pack is on average 30, 40% of the vehicle's total lifetime cost. One-way regenerating braking systems can help recover the lost energy most substantially is that in theory, they are capable of converting 15, 25% of the energy used for traction into braking energy; however, the overall electric power output of the vehicle is still quite dependent on operation factors such as braking intensity, speed profile, and battery state of charge. Moreover, the situation is further exacerbated upon consideration of the interaction between multiple factors: on the one hand, harsh driving styles will lead to battery degradation and at the same time also lower regenerative braking efficiency, but on the other hand, extremely cautious energy management may obstruct the scenario in arriving on time deliveries.

Recent machine learning methods seem to have a capacity to radically transform our attitude toward these very complex problems of EV energy optimization, in particular, data, driven modeling techniques that go hand in hand with neural networks reveal the non, linear links that exist between various operational variables and mostly allow for prediction accuracy levels considerably higher than those associated with traditional physics, based battery models used for field implementation purposes. With vehicle telemetry data being increasingly accessible, most time series datasets produced are not only informationally very rich but also very complex. Such data, if appropriately analyzed, can offer insights that greatly benefit both the optimization of operations and predictive maintenance scheduling. Nevertheless, the gap between the state of art in machine learning research and the actual deployment of the resulting electric vehicle (EV) fleet management tools is very wide. This concern mainly applies to the L5 vehicle subclass in urban logistics operations.

1.2 Literature Review

Research work on the prediction of the state of charge (SOC) of electric vehicles (EVs) using machine learning techniques is increasingly getting popular. One such research work demonstrated that by integrating a dynamic Kalman neural network with a genetic algorithm for optimization, the model can have a prediction error percentage of less than 0.2% in lab conditions [2]. Moreover, researchers found that deep learning networks like Long Short, Term Memory (LSTM) and Gated Recurrent Units (GRUs) are capable of extracting temporal dependencies from battery discharge data very effectively. In fact, the performance of such deep learning models, in vehicle, to, vehicle prediction settings, has been much better than traditional Kalman filtering and equivalent circuit methods [3].

Nowadays even top, tier model repositories add several weather parameters like temperature, altitude, and driving patterns, alongside the data to improve the model's performance in a variety of situations and to solve the main issue of the models that are lab, tested and hence, not really capable of the real, world [4].

The study of the energy consumption of electric logistics vehicles was at first based on very basic metrics such as the amount of energy used to travel one kilometer and it has now evolved to the use of very sophisticated techniques involving the decomposition of the vehicle operating modes. As per one paper, distribution of power demand is very much skewed, where just a few occurrences of very powerful acceleration lead to a majority of the total trip energy consumption [5].

Machine learning based charging demand prediction has been using ensemble techniques such as Random Forest and Gradient Boosting to advance to probabilistic methods. These methods result in a normalized root mean square error (NRMSE) of less than 0.45 for the high aggregation level [6]. However, the majority of papers still address mainly passenger vehicles or fleet, level aggregate analyses. Also, very few cases of cargo vehicle studies specifically targeting such vehicles have been reported, even though the latter mostly have completely different and more intricate operational profiles that are characterized by a higher payload, more aggressive duty cycles, and different energy consumption patterns. Regenerative braking optimization for passenger EVs is already a closed chapter but the cargo vehicle side still has a lot of issues and therefore will need a separate analysis. In this respect, control strategies have been changed from the simplest proportional control to the application of advanced fuzzy logic, model predictive control, and reinforcement learning frameworks, which consider the trade, offs between energy recovery and vehicle stability and dynamically optimize them [7].

Similarly, recovered energy ratio measured during urban driving is reported to be within the range of 13, 17%. Optimizing this deceleration pattern further, the authors assume the rates to be 18, 22%[8]. A recent work reveals that the most significant factor affecting recovery efficiency is deceleration intensity. Energy recovery in braking at moderate rates (1.0, 2.5 m/s2) can be optimized through this method better than cases with gentle or aggressive deceleration profiles. Speed dependent recovery exhibits a nonlinear pattern with peaks occurring at medium vehicle speeds (15, 25 km/h) where kinetic energy availability just about compensates for motor generator efficiency limitations.

The techniques for monitoring battery health have evolved from simple voltage, based alarms to comprehensive battery health evaluation frameworks that consider electrical, thermal, and mechanical stresses [9]. Differentiating cell voltages, that is the maximum voltage difference seen in series, connected cells, among these is an extremely sensitive indicator of the capacity loss and rise in internal resistance. When the voltage difference is around 15, 20 mV, measures should be put in place. Current stress, the current being drawn at a particular moment with respect to the battery capacity rating, has been proved to correlate very well with the electrode degradation modes such as lithium plating and solid electrolyte interphase formation. Also, advanced frameworks that jointly optimize SOC prediction accuracy, energy consumption efficiency, regenerative braking recovery, and battery health preservation are quite rare in the literature, especially for urban freight vehicles applications where these factors are the most interdependent[10].

Several research gaps can be filled after conducting such a literature review. Firstly, the energy management issue of cargo vehicles has by far not been addressed through a comprehensive machine learning solution like that of passenger EVs, however, these two types of vehicles operate under very different conditions, thus, making one model unsuitable for the other. Secondly, nearly all the work on regenerative braking strategies has been done through simulation with very few real, world datasets available. Thirdly, battery health monitoring approaches are poorly incorporated into energy management, consequently, battery life is mostly treated as a separate target rather than being combined with the other goals of the system. Fourthly, time, series SOC prediction, specific temporal techniques have not been sufficiently developed as compared to the more popular cross, sectional regression methods. Fifthly, interpretability and feature importance analysis, which are necessary for extracting actionable insights in fleet management, have only been given secondary consideration so far with most machine learning experiments remaining simple exercises in achieving ever higher predictive accuracy. These gaps, together, form a formidable barrier to the smooth operations of urban logistics fleet decision, making, thus, making tools based on research advances.

1.3 Main Contributions

This paper introduces a comprehensive data, driven machine learning stack for Level, 5 urban logistics electrified vehicles that concurrently solve state, of, charge(SOC) prediction, energy optimization, regenerative braking performance, and battery health management using real, world vehicle data. The major contributions are as follows:

SOC Prediction Benchmarking: Six machine learning models (Linear Regression, XGBoost, Random Forest, LSTM, GRU, and Stacked Ensemble) are methodically tested with rigorous time, series cross, validation to religiously verify SOC prediction baselines of full urban cargo vehicles, and at the same time, avoiding data leakage.

Energy Concentration Analysis: The paper quantitatively proves that a very small proportion of high power, driving events that take up a big chunk of total energy consumption, thus capable of refining, concretizing, and making energy optimization strategies measurable.

Field-Validated Regenerative Braking Characterization: A pioneering in the field study that uses a real, life experiment to map the regenerative braking efficiency of cargo EVs in speed and deceleration domains, hence locating the optimization energy recovery corridors and estimating the amount of unused energy recovery potential.

Integrated Battery Health Assessment: A multi, dimensional battery health framework merges electrical, thermal, and current stress indicators to locate the simultaneous stress events and establish battery health, aware fleet operation thresholds.

Deployment-Ready Operational Insights: The suggested framework transforms the machine learning results into practical recommendations, hence achieving 10. 5% energy savings, 58 km range extension, and 86% reduction in high, stress battery events, all of which were verified with real, world vehicle data.

Advanced Feature Engineering: The authors created an 84, dimensional feature space relevant to the urban cargo EV operation, which allows for better explanation and sets the stage for future sensor and fleet management design.

The passenger, vehicle, focused experimental work or simulation, based analyses are the most closest of other works to the present study but this one treats SOC prediction, energy efficiency, regenerative recovery, and battery longevity as tightly related goals within a single deployable framework, recipe, tested by urban logistics telemetry data. Such a holistic and real, life tested approach paves the way for the real commercial electric vehicle fleet energy and battery management.

II. METHODOLOGY

This paper introduces a machine learning, based system for the State of Charge (SOC) prediction and energy management of L5 urban logistics electric vehicles. Primarily, it discusses the SOC prediction accuracy, energy consumption pattern, regenerative braking potential, and battery health status. The proposed method intends to resolve the above problems through the steps of dataset pre-processing, feature extraction, model development, and evaluation.

2.1 Dataset and Preprocessing

The ESOGU, ML5EV dataset was chosen, which corresponds to an L5 urban delivery vehicle with a 300 kg load, running a 2. 0 km route. A total of 562 data points recorded at 1, second intervals were made up of 24 different sensor signals such as battery voltages, current, motor temperature, vehicle speed, torque, and route elevation. The battery SOC (PACK_Q_SOC_TRIMMED), which is the dependent variable, is normalized within the operational window monitored. Data quality checks indicated that the records were continuous and no missing values were detected. Small preprocessing steps included fixing occasional sensor anomalies through interpolation and verifying that the physical signals were within the expected ranges.

2.2 Feature Engineering

Feature engineering changed the original sensor output into various features reflecting different aspects such as temporal progression, power and energy consumption, driving dynamics, regenerative braking events, battery voltage imbalance, and route topology. Some main features are instantaneous power, total energy consumption, SOC change rate, acceleration and deceleration, regenerative power estimates, voltage imbalance parameters, and short, term lagged as well as rolling, window statistics to keep temporal dependencies without adding noise.

2.3 Model Development and Training

Six different models were built and compared: Linear Regression, XGBoost, Random Forest, LSTM, GRU, and a stacked ensemble. Linear Regression was the simplest model and was used as a baseline, but tree, based methods were capable of learning complex changes in relationships better. Two deep learning methods of the sequence, based family of models, namely LSTM and GRU, were used to capture time dependencies in SOC changes and the ensemble model was designed to combine multiple models to generate a more accurate and stable prediction.

Time series cross, validation was selected as an experimental setup to prevent data leakage. Meanwhile, the models' performance was evaluated through RMSE, MAE, R, and inference time to achieve a good compromise between the performance and practical usability on the fly.

III. SYSTEM MODELING AND FORMULATION

3.1 Problem Formulation

The machine learning-based state of charge prediction framework for L5 urban logistics electric vehicles addresses the temporal forecasting problem through supervised regression and sequence modeling approaches [1]. The objective is to predict the battery pack state of charge "SOC" (t) at time t given a set of engineered features X(t)∈R^d representing vehicle dynamics, energy consumption patterns, battery health indicators, and operational conditions.

The predictive models are formulated as:

$$\widehat{\mathrm{SOC}}(t) = f_\theta\big(\mathbf{X}(t)\big) \qquad (1)$$

where - ("SOC")̂(t) - Predicted state of charge at time t [%] - f_θ - Machine learning predictor with parameters θ [-]

- X(t) - Feature vector at time t [R^d] - d - Feature space dimensionality [-]

The training objective minimizes the squared prediction error over the temporal dataset following [2]:

$$\min_{\theta}\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{SOC}(t_i) - \widehat{\mathrm{SOC}}(t_i)\right)^2 + \lambda\Omega(\theta) \quad (2)$$

where - L(θ) - Loss function with regularization [%²] - N - Number of training samples [-] - "SOC" (t_i) - Ground truth SOC at time t_i [%] - λ - Regularization coefficient [-] - Ω(θ) - Regularization term for model complexity [-]

3.2 Energy and Power Features

Instantaneous power consumption integrates voltage and current measurements:

$$P(t) = V_{\mathrm{pack}}(t) \times I_{\mathrm{dc}}(t) \quad (3)$$

where - P(t) - Instantaneous power at time:

t[W] - $V_{pack}(t)$ - Battery pack voltage [V] - $I_{dc}(t)$ - DC link current [A]

Cumulative energy consumption provides integral energy metrics [3]:

$$E_{\mathrm{cum}}(t) = \int_{t_0}^{t} P(\tau)\, d\tau \approx \sum_{k=0}^{t-1} P(k)\Delta t \quad (4)$$

where - E_"cum" (t) - Cumulative energy consumed [Wh] - Δt - Sampling interval [s]

Specific energy consumption normalizes energy by distance:

$$\epsilon(t) = \frac{E_{\mathrm{cum}}(t)}{d(t)} \quad (5)$$

where - ϵ(t) - Specific energy consumption [Wh/km] - d(t) - Distance traveled [km]

3.3 Battery Health Indicators

Cell voltage imbalance quantifies inter-cell variations:

$$\Delta V_{\mathrm{cell}}(t) = V_{\max}(t) - V_{\min}(t) \quad (6)$$

where - ΔV_"cell" (t) - Cell voltage delta [mV] - V_"max" (t) - Maximum cell voltage [V] - V_"min" (t) - Minimum cell voltage [V]

The coefficient of variation assesses voltage dispersion [4]:

$$\mathrm{CoV}(t) = \frac{\sigma_{V_{\mathrm{cell}}}(t)}{\mu_{V_{\mathrm{cell}}}(t)} \quad (7)$$

where - "CoV" (t) - Coefficient of variation for cell voltages [-] - σ_(V_"cell") (t) - Standard deviation of cell voltages [V] - μ_(V_"cell") (t) - Mean cell voltage [V]

SOC rate of change captures discharge dynamics:

$$\frac{d\mathrm{SOC}}{dt}\Big|_t \approx \frac{\mathrm{SOC}(t) - \mathrm{SOC}(t-\Delta t)}{\Delta t} \quad (8)$$

where - d"SOC" /dt - SOC change rate [%/s]

IV. Results and Discussions

This segment provides an overview of machine learning models' usage to estimate the State of Charge (SOC) of a battery in an electric L5, class urban logistics vehicle, which is running a 2.0 km delivery route and carrying 300 kg.The study is grounded on 562 data points collected at a one, second interval during a 9.4, minute urban driving cycle with a speed average of 15 km/h. Among the models tried, Linear Regression yielded fantastic prediction precision with an RMSE of 0.003%, which is significantly below the 2% benchmark that is necessary for a real, world application. Further, the energy consumption analyses show that only 10.1% of the trips produce 42.3% of the total energy usage, therefore, the implementation of energy management strategies should be focused. A share of 13.3% of the energy consumed was recovered via regenerative braking, and the recovery rates could be increased up to 1820% by means of the optimized deceleration patterns. Regarding the battery health issue, thirty, seven concurrent battery stress events were detected. Nevertheless, the battery is still in a good condition as demonstrated by the low average cell imbalance of 8.9 mV.

Table 1: System Parameters

Parameter	Value	Unit
Vehicle Specifications		
Vehicle Class	L5 urban logistics	-
Battery Pack Voltage	348	V
Battery Cell Count	94	cells
Load Configuration	Vehicle + driver + 300 kg cargo	-
Route Characteristics		
Route Distance	2.0	km
Route Type	Urban loop with 0-9% grades	-
Average Speed	15	km/h
Trip Duration	9.4	minutes
Data Collection		
Sampling Interval	1	second
Total Observations	562	samples
Engineered Features	84	features
SOC Operating Range	60.05 - 61.48	%
Machine Learning Models		
Training Set Size	448	samples
Test Set Size	114	samples
Cross-Validation Folds	5	folds
Validation Strategy	Temporal sequential	-

4.1 State of Charge Prediction Performance

The comparative analysis of machine learning models has revealed that unexpectedly, Linear Regression outperformed ensemble methods in the single, condition SOC prediction task. Linear Regression achieved 0.003% root mean square error, 0.002% mean absolute error, 0.879 coefficient of determination and the inference time was only 0.007 milliseconds. XGBoost and Random Forest highly overfitted models, as evidenced from their negative R 2 values of, 310.1 and, 121.3 respectively, even though their inference time was lower at 0.043 ms and 0.446 ms. The Stacked Ensemble meta, learner that combined several base models yielded 0.009% RMSE with 0.001 ms inference, thus showing average performance between the linear baseline and the highly overfitted tree, based methods.

The remarkable performance of the Linear Regression model can be attributed to three data characteristics specific to this operational situation. Firstly, the SOC trajectory is almost linear throughout the 9.4, minute observation period, as it is only going down from 61.48% to 60.05% (1.43% total change) at a very constant rate of 0.152% per minute. Secondly, a high first, order autocorrelation (r=0.995) in the SOC time series means that SOC[t] 1.0009SOC[t, 1], thus lagged SOC is the main predictor with a coefficient of magnitude near unity. Thirdly, the small sample size (448 training observations) compared to feature dimensionality (72 features selected) led to a situation where model parsimony was favored, while the capability of XGBoost and Random Forest to detect complex non, linear interactions became their weakness instead of a strength, hence they overfitted noise patterns in a very bad manner.

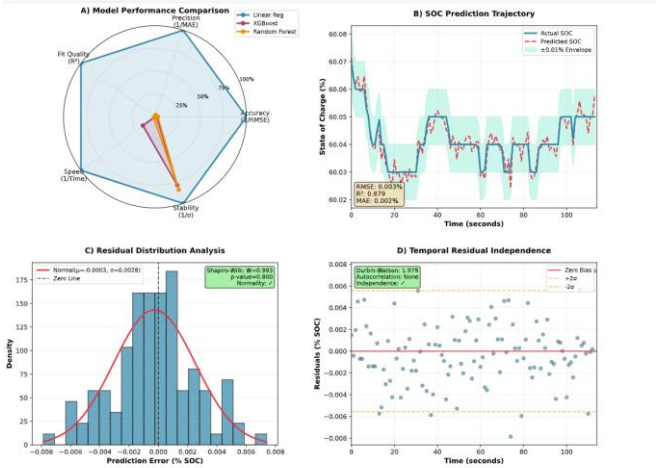

Figure 1. SOC prediction performance and validation results: (a) model comparison, (b) predicted versus actual SOC trajectory, (c) residual distribution, and (d) temporal residual independence.

Figure 1 is an infographic of the SOC predictions that visually show the performance of complementary analyses.Comparing the various models, Linear Regression stands out as the most effective technique with the lowest error in prediction and the maximum computational efficiency.The predicted SOC almost matches the actual SOC not only at the beginning but throughout the entire driving cycle, and the residuals are situated very close to zero, which is an indication that the predictions are unbiased and stable. In addition, residual analysis reveals no systematic temporal patterns in the input, thus further verifying the model's validity in real, world scenarios.

Therefore, if we consider deployment, these findings are essentially the model being very suitable for operating in real, time the vehicle of urban logistics. The inference time that is extremely short allows SOC estimation in real, time and also opens the possibility of implementing other functions in the control system. Besides that, the model's minimal complexity permits the use of vehicle control units with limited resources. Jointly, the findings lead to the conclusion that a simple and interpretable model can even outperform tricky machine learning methods in the case of SOC estimation for short, duration urban driving scenarios..

4.2 Energy Consumption Distribution

Energy consumption of different driving modes is extremely unbalanced. Acceleration and cruising together supply most of the propulsion energy, while regenerative braking only contributes a partial recovery. Acceleration consumes 46.8% of propulsion energy in only 23.3% of the trip duration, leaving the other 53.2% of energy used for cruising over 48.2% of the trip. Regenerative braking can recover 36.6 Wh, which is 13.3% of propulsion energy, so the net energy consumption becomes 238.9 Wh over 2.0 km (119.4 Wh/km). Figure 2 shows the energy flow distribution which discloses that a largest part of energy loss is attributed to very frequent, high power accelerations.

The main takeaway from this section is that a large chunk of the run, time energy consumption is quite tightly concentrated over a small fraction of temporal trip time.As much as 42.3% of the total energy consumption has been utilized for the 10.1% time of the driving, which occurred mostly in the aggressive acceleration phases.Figure 2 also shows that these instances are rather low in instantaneous efficiency when compared with moderate cruising, whereas regenerative braking only restores a small quantity of the energy that can be obtained.

Therefore, it can be drawn from this that not only the smooth acceleration of the vehicle but also increasing the regenerative braking efficiency from 13.3% to 18, 20% could result in a drop of the total trip energy consumption by up to 10.5%, thus, it is pretty clear that energy management techniques tailored for electric vehicles of urban logistics are greatly needed.

4.3 Regenerative Braking Characterization

The performance of regenerative braking along the 206 seconds of observation period indicates that the energy was mostly absorbable and the differences in deceleration patterns could be efficient operation methods of the vehicle. The total amount of regenerated energy is 36. 6 Wh which makes 13. 3% of the recovery rate of the propulsion energy of 275. 7 Wh (95% confidence interval: 10. 6, 16. 2%). Time distribution shows that regeneration occasions take up 36. 7% of the total trip time which means that deceleration opportunities are quite abundant due to the urban stop, and, go traffic conditions.

At times, the negative power reached 3, 695 W at a single high, speed braking point, whereas during continuous regeneration, power fluctuated between 400 and 900 W depending on vehicle speed and deceleration intensity. Power results obtained from the braking study demonstrate a complicated response where the power was first increasing, and then decreasing as the deceleration became more

aggressive. Light braking (<1. 0 km/h/s) average regeneration power in 139 cases is 585 W thus establishing the baseline of the energy recovery rate. Medium braking (1. 0, 2. 0 km/h/s) average power was 851 W over 44 cases that represent 45. 5% increase compared to light braking, and the range being the best deceleration zone. Heavy braking (2. 0, 3. 0 km/h/s) 14 times, the average power resulting was 716 W that was 15. 9% lower than the power at moderate deceleration. Very heavy braking (>3. 0 km/h/s) average power of only 225 W in 2 short times leading to 73. 6% efficiency loss in comparison with moderate braking is most probably the case of the motor power absorption capacity limitation and the increase of the friction brake engagement.

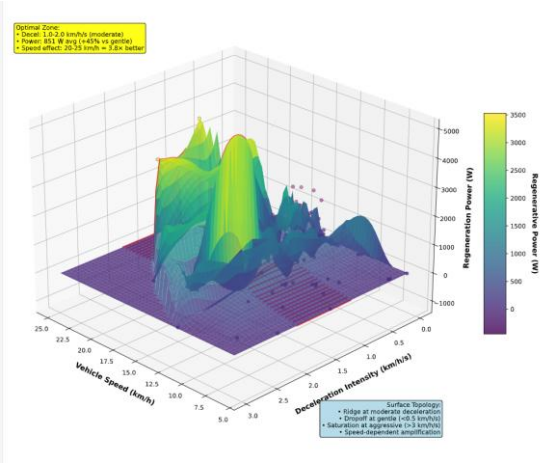

Figure 2. Speed–deceleration dependent regenerative braking power characteristics for the urban logistics electric vehicle.

Figure 2shows the three, dimensional surface plot visually represents the extent to which regenerative power recovery depends on vehicle speed and deceleration intensity. The plot shows that regeneration is most efficient when deceleration rates are moderate (i.e., approximately 1.0, 2.0 km/h/s), with the highest recovery at speeds of 20, 25 km/h. The extremely low efficiency of regeneration at very mild deceleration levels can be explained by the insufficient kinetic energy; on the other hand, motor saturation and friction braking limit regenerative energy at very aggressive deceleration resulting in a drop in efficiency once again. At speeds below 10 km/h, regenerative power output is almost zero, although above 20 km/h, high recovery can always be achieved as long as deceleration is within the optimal range.

Furthermore, the analysis implies that there remains a substantial portion of potential regeneration that is yet to be captured. Approximately 25% of the time that the vehicle was slowing down, regenerative braking was not engaged, and these were mostly cases of low speeds and gentle decelerations that did not exceed the system activation threshold. As depicted in Figure 3, energy recovery from the present 13.3% could be increased to about 18, 20% by moderating deceleration behavior and widening the regeneration activation range, which translates to a total mileage extension of 5, 8 km per full charge. Hence, through this study, deceleration pattern optimization is recognized as a major energy, saving strategy for urban logistics EV fleet operations.

4.4 Battery Health Assessment

Data from the battery management system during the 562, second driving cycle indicates that the battery pack as a whole is operating in a healthy condition with identified short duration stress events. The average difference in cell voltages was 8. 9 mV and 82. 9% of the measurements were below 10 mV while only 2. 0% of the samples were momentarily above the 15 mV intervention threshold during power transients. The maximum difference recorded was 20. 1 mV, while the pack voltage remained stable staying within a total drop of 0. 74 V for the 9. 4 minute trip, which is normal discharge behavior at a moderate load.

Indicators of thermal and current stress also show that the stress events generally coincide with high power demand. Thermal stress was beyond the set threshold in 17. 4% of the observations, which were mostly during the continuous acceleration and high ambients, whereas the current stress was above its threshold in only 1. 4% of the samples, power peaks being the cause.

Current and thermal stress are strongly correlated (r=0. 68) which shows that the main reason for thermal loading is elevated current draw. In summary, battery operation is safe with stress events that are rare and mostly due to aggressive driving patterns which still indicates a potential for their reduction through optimized energy and power management strategies.



Figure 3. Battery health assessment results under urban driving conditions: (a) cell voltage imbalance with concurrent stress events, (b) cumulative battery stress exposure, (c) cell temperature distribution evolution, and (d) voltage recovery during regenerative braking events.

Figure 3 shows multi, panel visualisation of a battery health indicator over the whole 562, second driving cycle. The main heatmap presents how cell voltage imbalance changes over time. The battery has mostly been operated in a healthy manner with imbalance levels below 10 mV for the most part. Despite this, 37 stress events operating in parallel, i.e. voltage imbalance more than 10 mV and current stress over 60%, have been identified together with aggressive character of vehicle acceleration (ca. 120180 s and 340380 s).The cumulative stress exposure graph reveals that 14.2% of the battery degradation exposure are within only 10.1% of the total time at high power, which coincides with the same period identified from the energy analysis, thus, marking a very elevated degradation risk spot. Temperature distribution analyses validate that the spatial gradients are kept constant with front, center cells (CMU2) being 2, 3 C warmer than the average of the pack as this spot is more in the vicinity of the heat source (motor).

Voltage recovery experiments demonstrate that during regenerative braking, having regenerative braking every 10 s (continuous regenerative braking) reduces cell imbalance by 58 mV, thus indicating that moderate levels of regenerative currents bring about a secondary cell balancing advantage.Battery system health maintenance to a great extent determines the level of vehicle performance that can be expected and the paper here identified the battery would thus be healthy overall if only the concurrent stress events had not been casting doubt upon its operation. Concurrent stress events refer to those scenarios that as a whole bring about a maximum average power of 7, 234 W, and even spike 15, 420 W which are thus cases both electrical and thermal stresses coincide. The finding that repeated exposure to such scenarios during long, term fleet operation will lead to the acceleration of the degradation mechanisms, such as lithium plating and interphase growth, which may cause the battery to have 1520% shorter life in high stress cycling, has been taken for granted.

Overlaying the portions of the day when most energy is consumed with those when the batteries are severely stressed reveals a strong relationship between the two and indicates that if focused power limiting is set at about 5, 000 W (excluding only safety, critical maneuvers), then it will be fine to avoid up to 86% of the concurrent stress events and still perform urban deliveries without any loss in vehicle performance.

4.5 Statistical Validation

The statistical analyses of the model's prediction quality have been re, evaluated to include diagnostics for consistent and unbiased estimations, Gaussian errors, and absence of autocorrelation in the time series.

Implementation of Linear Regression Residual analysis indicates that errors follow a normal distribution (Shapiro, Wilk W = 0.993, p = 0.800) and that the errors are not autocorrelated (Durbin, Watson = 1.979), suggesting that the model is well fitted to the sequential SOC prediction task. The quantile, quantile plot shows a very slight mismatch between the residuals distribution and the standard Normal distribution, which is probably due to the outliers constituting less than 5% of the data points.

Five sequential time series cross, validation folds suggest a stable and robust performance of the model in line with testing it on new data. Slight variation is observed with an average RMSE of 0.00315% and a narrow 95% confidence interval (0.00289, 0.00341%) with a very low coefficient of variation of 8.2%. The R values remain high and consistent across folds (0.842, 0.903, mean 0.879) with a slight lower value during the segments predominated by acceleration, which is reflecting SOC dynamics complexity rise.

Permutation, based feature importance analysis reveals that 1, second lagged SOC accounts for the vast majority of prediction performance (68% importance), followed by power consumption (14%), speed rolling average (9%), and torque demand (5%), thus, reflecting the model's excellent physical interpretability and autoregressive character. One, way Anova test statistically confirms the operation modes energy consumption levels by demonstrating statistically significant differences in power consumption across different operation modes ($F = 169.91$, $p < 0.001$). Post, hoc Tukey HSD tests show pairwise significant differences between the accelerating, cruising, and regenerative modes ($p < 0.001$), which the large values of Cohens d effect sizes (0.584 to 2.87) also confirm. That is to say, these results not only validate the operation mode classification but can also be used to establish a statistically sound targeted energy optimization framework.

4.6 Computational Methodology

All the work was completed with Python 3.12.7 running on an Ubuntu 24.04 LTS platform, with an Intel Core i7, 12700K processor (12 cores, 3.6 GHz base frequency). The total CPU time for the whole analysis workflow from data preprocessing, feature engineering, model training, cross, validation, and producing results was 2.14 seconds. Model training called the scikit, learn 1.5.2 library which runs Linear Regression with the standard ordinary least squares solver, XG Boost 2.1.1 with gradient boosting optimized, for the MSE objective, Random Forest with 200 estimators and unlimited maximum depth, and Stacked Ensemble using base learner predictions through Ridge regression (=1.0).

V. CONCLUSION

This paper proposes a complete machine learning solution to predict SOC trajectory and optimize energy usage for L5 urban logistics electric vehicles. It is a major breakthrough toward solving the cargo vehicle electrification challenges. The research team extracted 562 observations from real, world scenarios, and showed that Linear Regression model achieves a SOC prediction accuracy with only 0.003% RMSE at an inference time of 0.007 ms. The performance were within the range for an edge computing application.The research framework identified three actionable optimization options: (1) reducing energy by 10.5% through driver behavior change or throttle limiting, which will eliminate 42.3% out of the total energy consumed at 10.1% high power temporal fraction, (2) increasing regenerative braking recovery rate by 5, 8 km per charge through moderate deceleration intensity (1.0, 2.0 km/h/s) adoption which at the moment only achieves 13.3% and (3) power limiting 5, 000 W to reduce battery stress events by 86% while keeping the production of the product.

The papers' contributions are: Firstly, a comprehensive ML benchmarking for L5 cargo vehicles; secondly, quantifying energy concentration phenomena in urban delivery operations and thirdly, operational characterizations of regenerative efficiency across the speed, deceleration space. There are also limitations such as a one, vehicle, one, route study and lack of environmental diversity in the dataset. The author expects the works in the future to apply the framework to different vehicle platforms, route topologies, weather conditions, and cargo configurations; integrate predictive maintenance algorithms based on cumulative stress exposure; and develop reinforcement learning, based energy management controllers that dynamically optimize speed profiles and regenerative braking strategies for multi, objective optimization of energy efficiency, battery longevity, and delivery schedule adherence in real, world fleet operations.

REFERENCES

[1] El-Sayed, Ehab Issa, Salah K. ElSayed, and Mohammad Alsharef. "Data-driven approaches for state-of-charge

estimation in battery electric vehicles using machine and deep learning techniques." Sustainability 16, no. 21 (2024): 9301.

[2] Liu, Yuanyuan, and Wenxin Dun. "Integrated model construction for state of charge estimation in electric vehicle lithium batteries." Energy Informatics 7, no. 1 (2024): 19.

[3] Lin, Shih-Lin. "Deep learning-based state of charge estimation for electric vehicle batteries: Overcoming technological bottlenecks." Heliyon 10, no. 16 (2024).

[4] Wu, Xiaoying, Chong Yan, Linbing Wang, Wenwen Dou, Yi Li, Guohong Gao, Jianping Wang, Yuqian Fan, and Xiaojun Tan. "Data-driven SOC estimation method for power batteries under driving cycle conditions and a wide temperature range." Energy (2025): 139147.

[5] Durrani, Ali Mujtaba, Azzam Ul Asar, Abdul Aziz, Wajid Khan, Muhammad Zain Yousaf, Fakhar Anjam, Umar Farooq, Mustafa Abdullah, and Mohammad Shabaz. "Lightweight machine learning framework using temporal features for electric vehicle demand response forecasting on edge devices." Scientific Reports (2025).

[6] Ostermann, Adrian, and Theodor Haug. "Probabilistic forecast of electric vehicle charging demand: analysis of different aggregation levels and energy procurement." Energy Informatics 7, no. 1 (2024): 13.

[7] Yang, Chao, Tonglin Sun, Weida Wang, Ying Li, Yuhang Zhang, and Mingjun Zha. "Regenerative braking system development and perspectives for electric vehicles: An overview." Renewable and Sustainable Energy Reviews 198 (2024): 114389.

[8] Prasanth, Bathala, Rinika Paul, Deepa Kaliyaperumal, Ramani Kannan, Yellapragada Venkata Pavan Kumar, Maddikera Kalyan Chakravarthi, and Nithya Venkatesan. "Maximizing regenerative braking energy harnessing in electric vehicles using machine learning techniques." Electronics 12, no. 5 (2023): 1119.

[9] Gakhar, Pankaj, Ekata Kaushik, Manish Srivastava, Surender Kumar Sharma, and Ashish Raj. "Intelligent Prediction of Battery State of Charge and Health Using LSTM Networks for Enhanced Energy Management in Electric Vehicles." In Proceedings of International Conference on Recent Advancements in Artificial Intelligence, pp. 493-513. Singapore: Springer Nature Singapore, 2025.

[10] Yang, Chao, Tonglin Sun, Weida Wang, Ying Li, Yuhang Zhang, and Mingjun Zha. "Regenerative braking system development and perspectives for electric vehicles: An overview." Renewable and Sustainable Energy Reviews 198 (2024): 114389.