

A Bayesian-Optimized Deep Learning Framework for State of Charge and Remaining Useful Life Estimation of Lithium-Ion Batteries in Electric Vehicles

Bhavana RH¹, Muneshwara M S², Shivakumara T³, Heenakousar Khuddubai⁴

^{1,2}Department of CSE, BMS Institute of Technology and Management , Bengaluru, Karnataka India.

³Department of MCA, BMSIT&M, Yelahanka, Bengaluru, Karnataka India.

⁴ Department of CSE, Smt.Kamala and Sri.Venkappa M Agadi College of Engineering and Technology, Lakshmeshwar ,Gadag, Karnataka India

¹bhavanaritti@gmail.com , ²muneshwarams@bmsit.in, ³shivakumarat@bmsit.in,

⁴heenak25799@gmail.com

²<https://orcid.org/0000-0003-4714-4100> and ³<https://orcid.org/0000-0002-5213-0953>

Abstract: Electric vehicles (EVs) significantly depend on lithium-ion batteries with monitoring of the State of Charge (SOC) and estimation of the battery life time being crucial to safety, enhancing operating life, and confidence by the user. Is the first of its kind to present a joint framework of SOC estimation and battery health and Remaining Useful Life (RUL) forecast. The method combines modern deep neural networks, namely Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and two-dimensional Convolutional Neural Networks (CNN2D), which were all optimized along the Bayesian code structure and do not require trial-and-error, but perform optimally. Based on these, classical machine learning baselines, Linear Regression, Support Vector Machine (SVM), XGBoost are employed in benchmarking. The models are then trained using rich time-series data including voltage, current, temperature and engineered features so that the models can capture the short-term dynamics and long-term degradation trends. In addition to the SOC estimation, the framework unites capacity fade modeling, cycle life and calendar aging studies, helping create precise RUL estimates. Through experimental analysis, Bayesian-optimized deep learning models are found to outperform conventional approaches consistently in terms of both SOC accuracy and prediction of the lifespan of EV batteries, providing a robust, data-driven addition to the battery management system of an EV. This dual functionality contributes not only to efficient energy consumption, but enables proactive maintenance methodology, so as to minimize down-time and the costs of operation.

Keywords: State of Charge (SOC), Lithium-ion Battery, Bayesian Optimization, Deep Learning, LSTM, BiLSTM, GRU, CNN2D, Battery Degradation, Capacity Fade, Remaining Useful Life (RUL), Cycle Life prediction, Electric Vehicle (EV), Battery Management System (BMS).

1. INTRODUCTION

The life blood of contemporary electric vehicles (EVs) lithium-ion batteries have become commonplace in small urban vehicles as well as electric super cars. To run safely and effectively, two aspects of battery monitoring are vitally important to these vehicles: not only is it essential to know precisely how much is left, so-called **State of Charge (SOC)**, it is equally important to know how long before the battery is depleted, referred to as **Remaining Useful Life (RUL)**.

Proper estimation of the SOC can assist the driver to prevent any unexpected loss of power, the most favorable charging and discharging pattern, and energy management to improve range. In the same way, trustful lifespan prediction enables vehicle owners and fleet managers to schedule maintenance or battery replacement prior to the significant drop in performance, thus making it cheap and avoiding failures. Nonetheless, the estimation of the SOC and RUL is by no means easy. The behavior of batteries is affected by a great number of factors which include: any temperature changes, usage patterns, charging rates, and chemical aging. Classical methodologies, including the equivalent circuit model and the Kalman filter are subject to the risk of deforming in a controlled environment but failing at the complexities of reality. We address these issues in this project through a **data-driven approach** that merges deep learning techniques such as LSTM, BiLSTM, GRU and CNN2D with **Bayesian Optimization** to enable automatic tuning of the hyperparameters. This is combined will allow the models to learn complex patterns in time series data that exists in battery data including voltage, current and temperature, but has the benefit of circumventing the inefficiency of manual parameter selection. In addition to SOC estimation, our system has a battery degradation analysis pipeline that can simulate capacity fade, cycle life, and calendar aging, which would be capable of predicting future RUL with considerable levels of

confidence. This paper offers a realistic upgraded Battery Management Systems (BMS) in an EV by combining the functions of proper SOC tracking and lifespan prediction into one, intelligent device. The result is not only higher driving reliability and a battery that is safer, but also smarter maintenance planning, longer battery life and the greatest possible returns on investments in electric mobility solutions.

1.1. Introduction to Electric Vehicles and Battery Management

Electric vehicles (EVs) are quickly changing our perception of transportation as they represent a greener, silent and more efficient vehicle compared to petrol-run cars. The battery will be at the core of any EV, the most important power source which will dictate the specifics of how far the car can go, how fast it can speed up and how long it will last before requiring a charge or a replacement. There exist various technologies of batteries and **lithium-ion batteries** have been seen to be used as the preferred battery type since they have high energy density, are lightweight and slightly have a long life.

The efficiency and the safety of an EV however relies strongly on the effectiveness of monitoring the battery and managing it. In this case the **Battery Management System (BMS)** comes to play an essential role. The BMS is the brain of battery, continuously monitors the parameters such as voltage/current/temperature / **State of Charge (SOC)** to keep them in ideal condition. It also protects against such hazardous circumstances as overcharging, deep discharge or overheating. In addition to monitoring, day-to-day battery monitoring is now anticipated with modern BMS solutions that are asked to predict the **Remaining Useful Life (RUL)** of the battery in other words when the battery will be at the end of its serviceable life. This data can assist EV owners in determining when maintenance is scheduled, as well as reduce the risks of developing an unexpected breakdown and obtain the maximum result of their investments. As the use of EVs becomes increasingly common, there has never been a greater need in the battery management industry to have more intelligent, data-driven battery management and this is opening the doors to new powerful technologies such as deep learning and Bayesian optimization.

1.2. Significance of State of Charge (SOC) in EV Battery Operation

Battery state of charge, commonly abbreviated SOC or State of Charge, is the term to describe the electric vehicle battery fuel gauges. Just as when a gas tank indicator reminds you how much petrol is available on a car, SOC does the same to a battery, letting you know the amount of usable energy available on the battery at any given time. It is important that the driver and the battery management system of the vehicle knows the SOC correctly so as to make informed decisions. Drivers directly feel the effects of SOC information on the estimation of range that is useful in avoiding cases where the vehicle abruptly runs out of power. It also has a bearing on driving patterns and charging patterns as well, promoting practices that can extend battery health. In the case of the Battery Management System (BMS) accurate estimation of SOC would allow charging and discharging of the battery within safe limits, making sure it is not affected by overcharging or deep discharge which can severely cut down battery life.

Due to numerous things that happen to the lithium-ion batteries, such as temperature, load, aging, charge/discharge rates, etc. it is hard but imperative to approximate the SOC. One of the most critical parameters in the operation of an EV battery is a reliable SOC measurement because it allows one to maintain vehicle safety, enhances performance consistency, and enables the development of smart energy management strategies.

1.3. Challenges in Accurate SOC Estimation

- Older batteries can absorb less charge than they did as new so that measuring voltage alone or current alone do not tell the whole story any longer.
- Fluctuations in temperature, such as a cold morning in winter or a warm afternoon in the summer, will cause the battery to behave differently, which is why it is hard to monitor the status of the charge.
- You cannot execute the abovementioned action, such as suddenly accelerate, hard hit the brakes or quickly charge the battery, and then the behaviour of this device becomes complicated and unpredictable, making estimation more difficult.
- The voltage, current, and temperature measuring sensors are not fool-proof-they may be associated with tiny errors or noise that perplexes the system.
- The voltage inside the battery varies in a nonlinear fashion as it charges and discharges, thus with no easy formula as an illuminating pattern.
- There are other variables such as humidity or even the driving vibrations that can include a minor influence to battery readings.places limits on physically based or electrical circuit-based model systems that are able to adapt to such realities.

Finally, no matter how one estimates SOC, the method must be capable of working in real time and swiftly, hence deployable within the vehicle itself.

1.4. Role of Deep Learning in Battery Monitoring

The trend of deep learning has been a game-changer when it comes to batteries and their monitoring as well as management especially when it comes to electric vehicles. Unlike the more traditional approaches where simple formulas or reduced models are heavily used, deep learning will be able to learn deep patterns using the actual data in battery determinations -- such as voltage, current, and temperature over time. This implies that it can learn the nonlinear hour-to-hour and day-to-day dynamics, and keep track of the changing conditions, batteries undergo in their life which ensures confidence in SOC and lifetime estimation. Training deep learning models like LSTM, BiLSTM, and GRU on many observations allows the model to learn both the short term variability and the long term degradation of the battery and so the estimation of SOC and lifetime becomes more viable and accurate. They further respond well to varying patterns of use, temperatures and the effects of aging without the necessity of continually manipulating them manually. This is another major strength because deep learning is capable of refining itself overtime as more data is provided making the operation of battery management systems to get smarter. This will enable more accurate estimates of the charge remaining in a battery, allow batteries to operate safely, and also enable easier estimates of when the battery will require maintenance or replacement. In short, deep learning offers flexibility, adaptability, and strong predictive capabilities that will be essential in the next generation of battery monitoring systems in EVs.

1.5. Inspiration behind Bayesian Optimization of Hyperparameter Tuning

Hyperparameter selection in training deep learning models In training of deep learning models, it is pivotal to get suitable learning rate, the number of layers, numbers of batch dimensions among others to ensure that optimal results are obtained. Such parameters are referred to as hyperparameters. Nonetheless, it is not easy to determine the optimal combination. Conventionally, individuals attempt numerous values at random or in a grid which would consume much time and CPU capabilities. Here the Bayesian Optimization comes to shine. Rather than blindly fitting all the possible combinations it brainily learns based on what has worked previously to infer which hyperparameters are most likely to work best next time. Consider it to be a guided search which learns with each step, by taking what has been learned in the past and channeling into the most promising subareas.

The method is more resource- and time-efficient since it does not waste on unnecessary experiments and should be more accurate in selecting better model parameters compared to manual tuning or random search. In the case of deep learning models, in particular complex ones used in battery monitoring where the performance matters most, Bayesian Optimization comes in handy in ensuring that the models are run efficiently without idling.

1.6. Scope of the Study and Research Contributions

This paper aims to research a data-driven intelligent system to properly measure the State of Charge (SOC) of a lithium-ion battery, as well as determine the life-cycle of such batteries as they aid in the operation of an electric vehicle. It uses advanced deep learning approaches with Bayesian Optimization to automatically tune model parameters to achieve high accuracy and adaptability to diverse battery states or cycling behaviour.

The Scope Includes:

- Training deep learning models using time-series measurements such as voltage, current and temperature (that can capture both short-term transient behavior and long-term degradation) of the battery.
- The modeling of battery degradation due to capacity fade and cycle life analysis, and calendar aging, to predict Remaining Useful Life (RUL).
- Judged by comparison of deep learning models to classical machine learning methods in order to identify improvements and confirm effectiveness.
- To develop an effective model that could be embedded in Battery Management Systems (BMS) to facilitate working in real time and predictive maintenance.

RESEARCH CONTRIBUTIONS:

- Presenting a new method one that combines both SOC estimate and temporal battery lifespan in a single common way.

- Deploying Bayesian Optimization to explore the optimal hyperparameters and improve the accuracy of the model and decrease the time taken to train.
- Proving experimentally that deep learning algorithms such as LSTM, BiLSTM, GRU and CNN2D outperform more traditional ones in terms of SOC accuracy and lifetime prediction.
- Offering information on how data-driven methods may accommodate the effects of batteries aging in the real world and with changes in the environment to enhance reliability in electric cars.

1.7. Problem Statement and Objectives

Problem statement: Offer accurate real-time SOC estimation and a credible lifespan (RUL/capacity fade) forecasting system of the lithium-ion battery in EV application by means of information driven models.

OBJECTIVES:

1. Compare and build SOC estimation models (LSTM, BiLSTM, GRU and CNN2D) with classical baselines.
2. Use Bayesian Optimization to discover the optimal hyperparameters of deep models.
3. Establish an interdisciplinary lifespan forecasting pipeline -- capacity fade modeling, cycle life prediction and RUL forecasting.
4. Test in standard metrics (RMSE, MAE, maximum error) of SOC and RUL accuracy concerning the lifespan.
5. Present implementation details, take into account deployment (Flask interface), and give reproducible code.

2. LITERATURE SURVEY

The proper estimation of State of charge (SOC) of the lithium-ion batteries persists to be an important issue in improving the efficiency and trustworthiness of the Electric Vehicles (EVs) performance. During the last several years (2019-2024), recent studies based on advanced machine learning and deep learning approaches, hybrid models, and optimization methods have addressed these issues leading to a tremendous rise in estimation accuracy and battery health monitoring.

Among the most significant areas of advances has been in application of deep learning structure, notably Recurrent Neural Networks (RNN) specifically the Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). These models are very efficient in incorporating the temporal dependency and non linear nature that is characteristics of charge/discharge cycles through batteries. As an example, Chemali et al. [1] have shown that LSTM networks were able to discover the long-term dependencies in the battery data and led to an increase in the accuracy of predicting the SOC even under a challenging usage regime. Likewise, Xiao et al. [2] developed an ensemble optimisation model based on GRU, and the overfitting was mitigated as well, and the robustness of estimation in real-time SOC estimation improved.

Besides pure deep learning it has been found that hybrid methods with integration of data driven models and conventional models or filters may have enormous advantages. The example would be the introduction of adaptive SOC values where Wang et al. [3] brought together Kalman Filters and neural networks in order to speedily react to dynamic driving conditions and battery aging. In [4], Vedhanayaki and Indragandhi applied an Unscented Kalman Filter (UKF) and Coulomb Counting and the method was able to mitigate noise and nonlinear effects that the sensor designs impart on the estimation of SOC and consequently used this method to elevate the effectiveness of SOC estimation. Such hybrid approaches are an intermediate between interpretable models and more data-driven approaches.

Subsequently, the optimization of hyperparameters especially in the deep learning models is also an important theme in the literature as it plays a vital role in the performance of the deep learning models. Bayesian optimization has been a successful tool to be employed in tuning the model parameters and it does not require manual searches extensively. In Eleftheriadis et al. [5] Bayesian optimization was used to tune Bi-LSTM networks, and the results reported the significant performance gain on SOC estimation with tuned models over tuning procedures based on manual tuning. Such automatic tuning enables the use of very accurate and low-computational consumption models in battery management systems of EVs. Another reason why researchers urged the consideration of battery degradation and environmental aspects in estimating SOC could be found in the prevalence of this concern in the research. The processes of battery aging, including capacity fade, cycle life decrease and temperature changes, may have a considerable impact on the accuracy of estimation in the long-term perspective. Due to temperature effects, Liu and Liu [6] have suggested temperature-aware models of the SOH (State of Health) that use

the temperature to change SOC predictions. Meng et al. [7] added aging-awareness to the data-driven models to lengthen the life of the battery estimating the SOC level. The contributions highlight the need to have integrated models that are able to estimate SOC and forecast battery life together in order to carry out proactive maintenance.

Comparisons also shed more light on the strengths as well as weaknesses of different machine learning algorithms in estimating SOC. Indicatively, a study by Khawaja et al. [8] addresses the problem by reviewing several methods and they concluded that ensemble-based machine learning algorithms provided a trade-off between reliability and computational speed. Ben Sassi et al. [9] compared ANN and the Kalman filter and stated that ANNs are more adaptive to unpredictable real-life conditions of battery operation.

In addition to model accuracy, has been that of real-time applicability and updating parameters online. Qian et al. [10] proposed adaptation of SOC estimation using dual Extended Kalman Filters and optimization algorithms that respond to the changing batteries dynamics. This method allows learning new data continuously during the usage of vehicles and creates a system that would not be susceptible to abrupt environmental or load changes. More recently, there have been works suggesting new hybrid architectures that comprise convolutional and recurrent layers, including hybrid CNN-GRU [11], to learn both spatial and temporal features of the data coming out of the battery. The others have incorporated stochastic optimization and momentum-based learning to enhance convergence of training and minimization of errors in predictions [12]. These developments reflect not only a blistering pace of development of a variety of estimation techniques in SOC but also a balancing of theoretical and applied requirements.

To sum it up, the developing trend in the reviewed literature is the transition to more intelligent, adaptable and mixed mode SOC estimation techniques that integrate physical knowledge with effective data-driven techniques. The inclusion of Bayesian optimization in hyperparameter optimization, consideration of the degradation impacts on battery management, and the in-real-time adaptive application of the model are the current technological curve in the study of battery management. It is based on these premises that our research collaborators have developed a Bayesian optimized deep learning framework to simultaneously estimate SOC and the life cycle of a battery, to improve the battery system performance of EVs greatly.

2.1. Review of Existing SOC Estimation Methods

The accurate estimation of the State of Charge (SOC) of lithium-ion batteries forms a key control activity in the proper management of an electric vehicle, yet it is not an easy component because over time batteries behave in a complex way. Conventional solutions such as Coulomb Counting and Kalman Filters claim to be estimating SOC using mathematical models and sensor data, but tend to have problems with battery deterioration and non-linearities. Machine learning, particularly deep learning, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) deep learning models, have been used by the researchers in order to address these problems in a more accurate manner and especially to tackle the time-dependency of the battery data. The combination of these deep learning models with classic filtering approaches has resulted in a more robust form e.g. hybrid approaches that respond better to evolving battery conditions and degradation. In addition, with the new technologies, such as Bayesian Optimization, such models become tuned automatically, performing better and not requiring any tedious manual adjustment. And there is also the realization that it is also important to consider factors such as temperature and battery wear in order to maintain SOC estimations to be accurate at all points during the battery life. Despite these advances, including ones that are able to make real-time solutions, gaps remain in making solutions fast enough to be used in real time and able to predict battery health in order to plan maintenance more effectively. All in all, the study opens up to smarter, more versatile methods of SOC estimation allowing to balance accuracy, robustness, and usability in EVs.

2.2. Overview of Deep Learning and ML Approaches

Lithium-ion battery State of Charge (SOC) estimation in electric vehicles has been completely transformed, and the approach with the use of deep learning and machine learning (ML) methods is a reality. Unlike conventional model-centered approaches, ML approaches can extract patterns directly in data, which explains why they perform so well in identifying the nonlinear and time-varying behavior of batteries. Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Random Forests are some of the models that have been used with some level of success. These newer architectures (particularly the Recurrent Neural Networks (rnn), such as Long Short-Term Memory (LSTM) and Gated Recurrent

Units (GRU)) have become more popular because of their abilities to learn about sequential data and long-term dependencies present in the charge-discharge cycle of a battery. Furthermore, there is also integration between CNN and recurrent neural networks, namely CNN-LSTM or CNN-GRU, and have been considered to extract spatial and temporal information in battery datasets. These data-driven methods do not only enhance the accuracy of SOC estimation but also provide flexibility to a variety of battery states and aging. Furthermore, hyperparameters are increasingly optimized using some form of automated algorithm such as Bayesian Optimization in order to further refine the models performance without trial and error. In general, deep learning and ML models provide an effective set of tools allowing one to create accurate, reliable, and scalable solutions of SOC assessment.

This is due to the innovative prepotency in deep learning that has transformed numerous sectors and State of Charge (SOC) of lithium-ion batteries is not an exception. In contrast with traditional algorithms based on pure mathematical models or even concrete physical battery models, data-driven modeling including deep learning and machine learning can capture the highly complicated, highly nonlinear, and dynamical characteristics of batteries in realistic conditions.

- **SOE Long Short-Term Memory (LSTM)** networks work well in SOC estimation since they learn and retain information which can be stretched to a long time. Such a capability is suitable to reproduce the temporal dependencies of the battery dynamics such as voltage recovery and hysteresis factors.
- **Gated Recurrent Units (GRU)** are another version of LSTMs though simpler but equally effective. GRUs can be trained significantly faster with fewer gates and parameters and take up less computational resources, thus being appealing to use in real-time continuous battery monitoring in electric cars.
- **Bidirectional LSTMs (Bi-LSTM)** are models a further stepping forward in terms of temporal modeling as they process the input and produce the output as sequences going forward as well as backward. This gives the network the chance to exploit past and upcoming context in the same way, useful when the behavior of the battery is determined by the predicted behavior.
- **Convolutional Neural Networks (CNNs)** and specifically 2D CNNs have demonstrated good results and performance in SOC estimation by processing multivariate time-series data (voltage, current, temperature). CNNs are capable of learning deep features and noise removal by use of spatial filtering (e.g. gridding) which deviates predictive accuracy. However, compared to RNN-based models, CNNs do not suffer so much of the vanishing gradient problem, and they are also easier to parallelize, therefore suitable in embedded and edge computing systems.

Alongside deep learning, traditional machine learning algorithms remain valuable for SOC estimation, offering simpler models and useful baselines:

- **Linear Regression** provides a straightforward, computationally inexpensive way to model relationships between inputs and SOC. However, it cannot capture the nonlinear complexities present in battery behavior, especially under varied operating conditions.
- **Support Vector Machines (SVM)** leverage kernel functions to map inputs into higher-dimensional spaces, allowing more detailed relationship modeling. Their performance heavily depends on kernel choice and parameter tuning, and they may struggle with long sequences or real-time SOC data.
- **Extreme Gradient Boosting (XGBoost)** is an advanced ensemble method on the basis of gradient-boosted decision tree which is highly valued due to its quality of accuracy and modeling nonlinearities and feature interactions. It usually performs better than the simpler designs, although it must be meticulously calibrated to explain the variations in battery chemistry, variations in manufacturing, and environmental influences.

On the contrary, ionic techniques such as Coulomb Counting are prone to errors over time, especially when discharging loads at high rates and in irregular use as it is often with EVs. Theoretically, Kalman Filters rely on the proper model of the noise and the assumptions which might be underserved in the case of aging batteries or unstable conditions, and the performance will be worse. Computational complexity is another facet that is worth considering. In order to find useful applications of SOC estimation techniques in electric vehicles, the algorithms need to be lightweight enough to be deployed in real-time on embedded computer hardware. Sadly, as classical algorithms seek to increase accuracy, they tend to heighten their computational requirements, this relationship forms a trade off that aims at hindering real time implementation.

2.3. Bayesian Optimization in Machine Learning Applications

Bayesian Optimization has proved to be a valuable asset in the field of machine learning that can effectively expedite the process of tuning hyperparameters of machine learning models at scale, and this

is especially the case with more complex and computationally expensive options such as deep neural networks. In comparison to more conventional grid or random search, whose scales can be time consuming and inefficient, Bayesian Optimization improves the exploration of the hyperparameter space through probabilistic model. It constructs surrogate function to estimate objective function and then chooses the most promising hyperparameters by doing a trade-off between the exploration (visiting new regions) and exploitation (improving known good regions). The process then allows one to converge more rapidly to optimal or near-optimal settings using reduced evaluations. The Bayesian Optimization is of particular benefit in scenarios where using a model based on deep learning, e.g., LSTM or GRU would result in a model that needs careful parameter tuning of learning rates, layer sizes, and regularization parameters, e.g., the State of Charge (SOC) of a lithium-ion battery state is estimated with Bayesian Optimization. Its automated and systematic process involves a reduced manual effort and expertise required which enables researchers and engineers to build more accurate and sound models with a lot of efficiency. In general, Bayesian Optimization is an effective and feasible method to improve machine learning processes, performance in areas that require accuracy and dependable predictions in particular.

3. SYSTEM REQUIREMENT SPESIFICATION

System Requirements Specification, is essentially a step by step delivery plan outlining what the system must do, and also, how it should operated. It is a valuable document as it assists developers, testers and the rest of the involved individuals to align concerning the objectives of the system before any construction efforts take place.

In our Battery State of Charge (SOC) Estimation and Lifespan Prediction system in an electric vehicle, the SRS makes it abundantly clear on what data it will use in the system, how the data will be processed by a system, what should be the results of the system and what kind of environments the system must work in. In this manner, we ensure that the system we are building will correctly report the amount of charge remaining in the battery as well as how far the battery will take us. A clear SRS will also prevent confusion and ensure that the project is not diverted in the wrong direction, causing the end product not to be what was expected.

3.1. Dataset Description and Requirements

The relevance and quality of the dataset is vital in the development of a good State of Charge (SOC) estimating and battery lifespan predicting system. In this project, the data will be a time-series data consisting of measurements taken on lithium-ion batteries in electric cars. The most important parameters are voltage, current, temperature and other calculated parameters are the mean values of voltage and current. Such data values are necessary due to capturing the dynamic nature of the battery when operated under various scenarios such as charging, discharging and resting. The data should be wide-ranging to provide diverse usage conditions, battery health conditions, and environmental conditions so that the model can be generalizable. Also, preprocessing was recommended to work with missing values, noise, and inconsistencies of the data, keeping the accuracy. It would also be important to have proper normalization or scaling to help the machine learning models learn. On the whole, the dataset is the basis on which the models are trained, validated, and tested, and its quality directly influences the results of the real-life application of the system.

3.1.1 Panasonic Battery Dataset Details

The Panasonic 18650PF lithium-ion battery dataset has been used and proven to produce high-quality results with high reliability because of which it is a great source of SOC estimation studies. This dataset carries Voltage, current and temperature serial measurements taken at high resolution over a number of charge and discharge cycles with different load profiles and at different environmental conditions. Every cycle is timed, and the synchronization can be presented, as well as the realistic dynamic behaviour of the battery within a real range can be modelled. Such a rich dataset would allow the researcher to study the battery performance degradation and compound models that could successfully recognize the trends of SOC under variable and constant current loads.

3.1.2 Data Preprocessing Techniques

Pre-processing plays a pivotal role of transforming crude sensor data into a clean usable form so that it can be used to train the model. The first is handling of incomplete data or corrupted data points through either imputing the missing data or excluding tainted records. Methods to reduce noise in current measurement, and voltage measurement include low-pass filtration and moving average filtration. Much more advanced signal denoising packages can also be used, e.g. Savitzky-Golay or Gaussian filters to maintain the integrity of the signal mean and variance whilst mitigating the amount of noise.

3.1.3 Normalization and Feature Selection

Normalization makes all of the features of the input within a similar scale range, thus avoiding bias when training the model. Common methods are Min-Max scaling to transform the values to a range of 0-1, or normalization to Z-scores, to standardise features against a Gaussian distribution. Correlation analysis, mutual information, and other feature selection methods can be used to select the key predictors in the estimation of SOC. Derived quantities current derivatives, average current over a period and charging and discharging time are also useful since they can further understand the implications of the electrochemical behavior of the battery making the model better informed.

3.2. Evaluation Metrics

To ascertain the reliability of the estimation models of State of Charge (SOC) and attain high accuracy of the models in reality, evaluation of the performance is important. Numerous quantities are used to measure the quantitative correspondence between the model predictions and the actual battery behaviour:

- **Root Mean Square Error (RMSE):** RMS determines the square root of the mean square of differences dropped between the predicted and actual SOC values. It provides a clear notion of how inaccurate the average prediction error of the model would be and lower the less the model would be accurate. RMSE is very sensitive to individual outliers, so this is a suitable measure to pick up large deviations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

n = total number of predictions

y_i = actual (true) value at index i \hat{y}_i = predicted value at index i

$(y_i - \hat{y}_i)^2$ = squared error for each prediction

- **Mean absolute Error MAE:** the mean absolute error (MAE) calculates the mean absolute error between predicted and real SOC. It gives a simple to interpret measure of the magnitude of average error, providing information on how far, on average, the predictions are of true values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **Maximum Absolute Error (Max Error):** This is the maximum absolute error where the single largest difference between the predicted SOC by the model and the actual SOC is recorded. It is notable in terms of knowing the worst-case prediction error which is essential to the safety-critical battery management systems.

$$\text{Max Error} = \max_{1 \leq i \leq n} |y_i - \hat{y}_i|$$

- **Coefficient of Determination (R 2 Score):** R 2 measures how much of the variance that exists in the actual SOC data that the model explains. Higher values nearer to 1, depict that they have a powerful correlation and higher quality prediction.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- **Computational Efficiency:** This is not a common error metric but the measurement of the training and inference times is crucial to guarantee the model can run in real time or near real time conditions that electric vehicle battery management usually has to deal with. A combination of the metrics offers a complete picture of what benefits and flaws the model has, and further additions will make it more accurate, and the predictions relating to SOC and battery lifetime will be reliable.

3.3. Safety and Reliability Considerations in Battery Management

Lithium-ion batteries in electric cars require security and soundness because of the prominent role that batteries take in vehicle execution and traveling security. Battery Management Systems (BMS) are developed to check not only the State of Charge (SOC), but also BMS-specific protection against conditions likely to cause damage, shortened lifetime, and even hazardous conditions like thermal runaway.

Key Safety Considerations Include:

- **Overcharge and Overdischarge Protection:** Too much charging of battery or a complete discharge may permanently damage the battery. These extremes should be prevented by the accurate monitoring of SOC with safe operating limits by the BMS.
- **Thermal Management:** Batteries are extremely susceptible to temperature. Battery materials are degradable by overheating and can heat up to burn. Critical in order to keep the battery in good health and avoid accidents is reliable temperature monitoring and control.
- **Cell Balancing:** Difference in cell voltages may also cause unequal aging or breakdown. Balancing makes sure that there are even charging currents distributed among cells, increasing battery life and safety.
- **Fault Detection and Diagnostics:** Abnormalities like internal short, loss of capacity or sensor failures are detected at an early stage to prevent major breakdowns and avoid maintenance at the right time.
- **Resistance to Environmental and Operational Changes:** The BMS must be highly robust to changing conditions, varying loads, temperatures and aging, so that SOC estimation and safety are monitored proportionately across all system challenges over the battery lifetime. Incorporation of current estimation techniques such as Bayesian-optimized deep learning models in the modern BMS can increase not only the levels of SOC estimation accuracy but also resonate the capacity of the system to detect probable failures proactively to perform safety related functions and lengthen battery life.

4. SYSTEM DESIGN AND ARCHITECTURE

The system optimizes battery information and gets the correct estimation of the remaining charge. It begins by gathering vital data such as voltage, current, and temperature of the battery and scrubs and polishes it to be ready to be analyzed. Smart algorithms of various types (including deep learning models) collaborate to accurately guess the state of charge of the battery. To ensure even more accurate predictions, Bayesian Optimization auto-adjusts the settings of the system. The system is able to provide the users with the results in simple to interpretable graphs, and all of the above can easily share its predictions with the rest of the applications via simple APIs. The design itself involves quick, consistent, and maintainable, hence renders itself to real electric vehicles. In general, it is designed to deliver credible information about the battery which allows ensuring the electric cars operate in the normal manner.

4.1 Requirement Analysis

The primary objective of the SOC estimation system is to fulfill the practical requirements of the electric vehicle application and needs as well as the expectations to the application displayed durably in the real environment. On a functional aspect, the system should easily process various streams of input data of the battery i.e., voltage, current, temperature, and capacity and through proper machine learning and deep learning methodologies, to optimize, which should further be enhanced with the use of Bayesian Optimization, it should correctly estimate the State of Charge (SOC) of the battery. Other than prediction, the system must facilitate the visualisation of results easily and use APIs to allow deployment into the system to enable it to be integrated with the existing Battery Management Systems (BMS).

On the non-functional side, the system is designed to provide timely predictions with efficient use of computational resources, scale well to larger datasets or additional sensors, and maintain robust security, especially when connected to cloud-based platforms. It should be maintainable, reliable under different operating conditions, and user-friendly, allowing engineers and researchers to easily interpret outputs and performance metrics. The system must also work smoothly with popular software frameworks like Python, TensorFlow, and optimization libraries, while running efficiently within typical hardware capabilities.

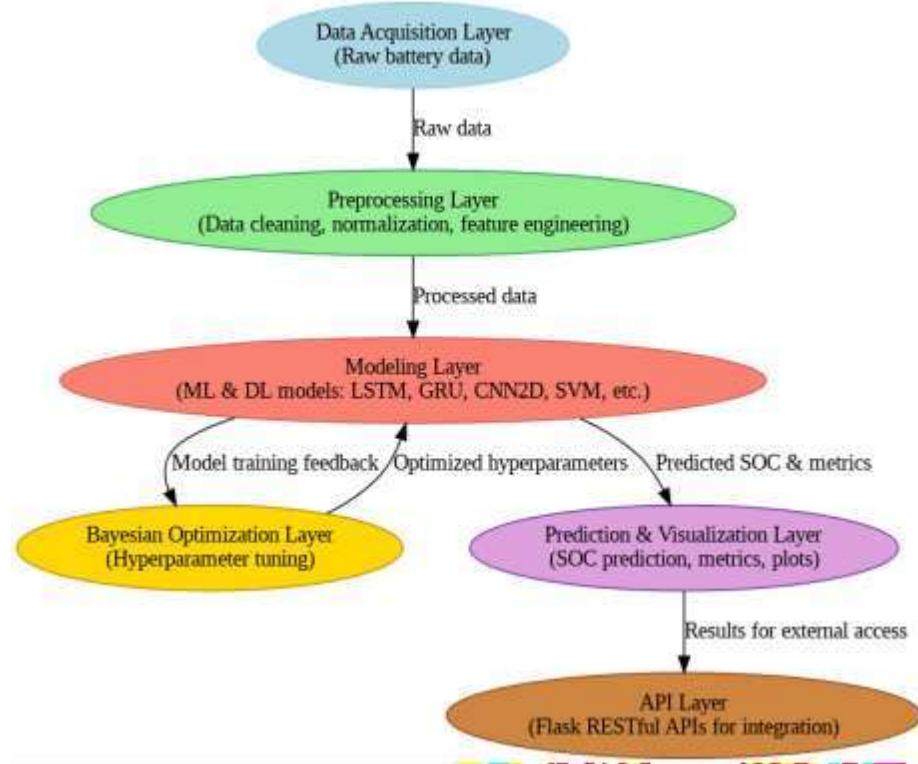


Fig 1: System Architecture Diagram

The Figure 1 shows the system architecture consists of several key modules working together to deliver reliable SOC estimation:

- **Data Acquisition Layer:** Captures raw battery data—voltage, current, temperature, capacity—from either the Panasonic dataset or live sensors.
- **Preprocessing Layer:** Cleans the data, applies normalization, and engineers features to prepare inputs for modeling.
- **Modeling Layer:** Hosts various machine learning and deep learning algorithms (LSTM, GRU, Bi-LSTM, CNN2D, SVM, Linear Regression, XGBoost) to estimate SOC.
- **Bayesian Optimization Layer:** Fine-tunes hyperparameters such as learning rate and batch size to improve model accuracy.
- **Prediction and Visualization Layer:** Generates SOC predictions and evaluation metrics, displaying them via graphs and dashboards.
- **API Layer:** Contains RESTful APIs implemented as Flask, with which the SOC predictions can be be interfaced with external Battery Management Systems or other applications.

4.2 Model Design Overview

4.2.1 Design of LSTM, GRU, and Bi-LSTM Models

The system is designed in a modular fashion. The Data Acquisition Layer is first, which acquires raw sensor data readings on the battery. This data is in turn preprocessed in Preprocessing Layer, where it is cleaned, normalized, feature-engineered in order to use it to train models. The Modeling Layer executes series of algorithms, such as complex deep learning predictors, LSTM, GRU, and CNN2D along with common machine learning predictors, SVM, XGBoost, and Linear Regression to forecast SOC. Bayesian Optimization Layer The Bayesian Optimization Layer is used to tune hyperparameters in the model dynamically in order to maximize accuracy in the predictions. Lastly, models and prediction, together with error metrics are visualized and made accessible through the API Layer, which enables them to be queried in real time with external systems such as EV dashboards.

4.2.2 CNN2D Architecture for Feature Extraction

The CNN2D model is important because any substantial feature created by the battery statistics is automatically extracted without any manual participation. Voltage, current, temperature and capacity inputs are formatted in a matrix format. CNN2D uses filters that identify local trends, such as volatility or repeating patterns, or trends, which are significant measurements of the behavior of batteries. Noise and dimensionality are minimized through repeated convolution and pooling layers such that only every subtle signal can be learned by the model, which cannot be learned by the conventional methods. To

increase prediction accuracy and robustness, the outputs of CNN2D can be combined with temporal models such as LSTM or GRU and such an approach can be made applicable to different battery chemistries and real-world operating conditions of electric vehicles.

Key CNN2D layers include:

- **Input Layer:** Receives the battery data inputs.
- **Convolutional Layers:** Detect local patterns using filters and ReLU activation.
- **MaxPooling Layers:** Reduce data size and filter noise.
- **Flatten Layer:** Converts multi-dimensional data into a 1D array.
- **Dense Layer:** Learns high-level abstract features.
- **Output Layer:** Produces the SOC estimate.

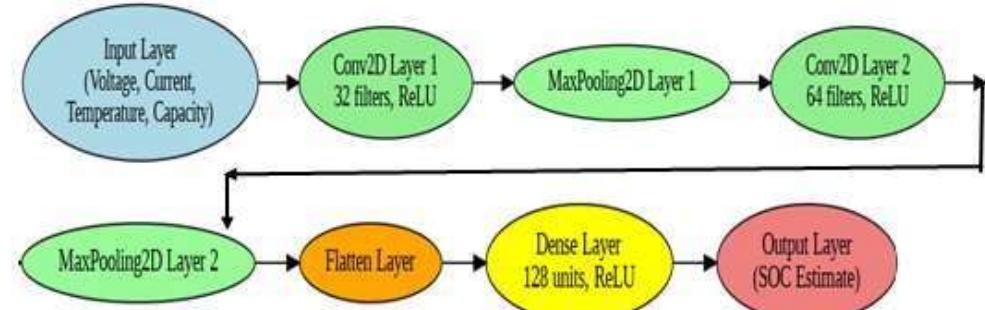


Fig 2: CNN2D Network Architecture Diagram

4.2.3 Bayesian Optimization Module for Hyperparameter Tuning

Such a module simplifies the process of optimizing hyperparameters in the models, by:

- Based on crucial battery functions such as voltage, current, temperature, capacity, average voltage and average current.
- Preprocessing data to ensure consistency.
- Employing deep learning models (CNN2D, LSTM, GRU) to learn battery behavior.
- Applying Bayesian Optimization to tune parameters such as learning rate, batch size, and network architecture.
- Producing an optimized model that generalizes well to new data.
- Delivering accurate SOC predictions expressed as battery charge percentages.

4.3 Validation of Design through Simulation and Theoretical Analysis

To confirm the reliability of the Bayesian-optimized deep learning models, extensive simulations and theoretical analyses were carried out. Models were trained using real-world battery data—voltage, current, temperature, capacity—and tested on unseen samples. Evaluation metrics like RMSE and Maximum Error demonstrated that models tuned with Bayesian Optimization consistently outperformed those tuned manually. Theoretical analysis further supported the model's stability, sensitivity to changing battery conditions, and the efficiency of the hyperparameter optimization approach. Overall, these results validate that the proposed system provides precise and reliable SOC estimates, effectively managing the complex and nonlinear behaviours of lithium-ion batteries in electric vehicles, contributing to safer and more efficient battery usage.

5. SYSTEM METHODOLOGY

To start the process, information is measured about the lithium-ion batteries, like the voltage, current, temperature and the charge actually being charged/discharged. Once this raw data is retrieved it is then introduced into a pre processing stage where the data is cleaned up and normalized. This helps by eliminating noise and the other advantage has been experienced where all the input features have gotten on the same order such that the models find it easy to learn the data. A variety of models of both machine learning and deep learning may be trained once the data has been prepared; these include complex neural networks like LSTM, BiLSTM, GRU, CNN2D and the more simple logistic regression, SVM and XGBoost. The insertion of Bayesian Optimization into these models creates a better rate of accuracy and performance. It is a smart procedure to optimally tune and therefore the models can be learned with a lesser degree of failed attempts and without the manual experimentation that is required. Once the most suitable parameters were discovered and the model was trained, one should test its performance based on standard accuracy metrics in order to ensure that it attains good reproducibility. The final model is thereafter implemented and deployed with the aid of a straightforward web program utilising Flask.

5.1. Research Methodology and Workflow

The research follows a well-organized and logical approach to estimate the State of Charge (SOC) of batteries in electric vehicles. It begins by collecting real-time battery parameters such as voltage, current, temperature, and the actual SOC. This raw data undergoes careful preprocessing, where noise is removed, missing values are filled, and data is standardized to make it suitable for analysis. The cleaned dataset is then divided into training, validation, and testing sets to ensure the models are properly trained and fairly evaluated.

Various models of machine learning and deep learning such as LSTM, GRU, BiLSTM, CNN2D and other well-known models such as SVM and XGBoost are built and trained on this data. Bayesian Optimization is used to automatically tune hyper-parameters (instead of using the time-intensive, less reliable, manual trial and error approach) and adjust the model settings effectively. The standard scorers like the RMSE, MAE, and accuracy are used to choose the measurement of the best-performing model. After deciding on the best-performing model, the model is implemented into a user-friendly web portal based on Flask. This web portal permits the user to feed new battery parameters and get real-time predictions of the SOC making this solution practical by implementing the solution in electric vehicle battery management systems. On the one hand, the methodology is focused on high accuracy and usability: It begins with rich data learning under different operating regimes, cleans the data, and searches a wide space of computation models--ranging from and using simple decision trees to large neural networks able to model complex time-varying phenomena. The automated hyperparameter tuning, in addition, boosts the predictive power of the model itself, and the last implementation in web-based form ensures the convenient access and use.

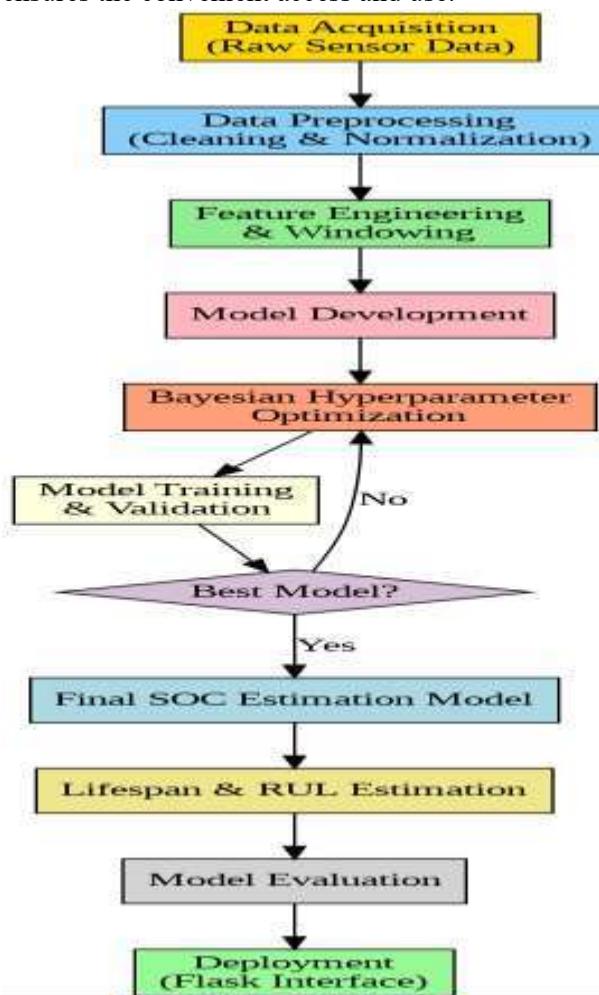


Fig 3: System Methodology

5.1.1. Dataset Splitting (Training, Validation, Testing)

In order to get a strong and standable model when estimating the battery SOC, the data set is properly separated as to form a set of a training, validation and testing sets. The model is trained to learn the patterns into the data using the training set. Throughout training the validation set is used to give feedback to adjust parameters of the model as a method of preventing over fitting utilizing the

performance of the model on previously unseen data. Once the training stage is reached, the test set is employed to determine the accuracy of the obtained final model and its capacity to apply to totally new data. Such orderly division of the dataset makes sure that the model not only achieves high accuracy on the data that it has been trained on but also behaves well in a real-life situation.

Key points:

- **Training set:** Used to train the model by learning data patterns.
- **Validation set:** Helps tune hyperparameters and avoid overfitting during training.
- **Testing set:** Used after training to objectively evaluate model performance on new data.
- Ensures the model generalizes well to unseen real-world data.
- Improves the reliability and accuracy of SOC predictions.

This architecture prevents the model to only memorize the data and not learning from it. It increases confidence that the motions estimates will be accurate when used in real-position, EV applications.

• **Data set**

Table 1: Data set

	Voltage	Current	Temperature	Capacity	Voltage_Average	Current_Average
0	4.20007	2.10781	12.053261	-0.14733	4.029908	-0.798057
1	4.18720	1.82851	12.053261	-0.14685	4.029996	-0.794400
2	4.08280	-0.03756	12.053261	-0.14686	4.029763	-0.795630
3	4.09519	0.21887	12.053261	-0.14681	4.029553	-0.796318
4	4.05835	-0.40586	12.053261	-0.14691	4.029270	-0.798233
...
8394	3.39904	0.00000	12.678676	-2.37330	3.356864	-0.338684
8395	3.39904	0.00000	12.458628	-2.37330	3.356626	-0.343039
8396	3.39904	0.00000	12.678676	-2.37330	3.356623	-0.342570
8397	3.39904	0.00000	12.458628	-2.37330	3.356849	-0.338466
8398	3.39904	0.00000	12.470211	-2.37330	3.357206	-0.332841

5.1.2 Justification for Deep Learning over Classical Methods

Well-known machine learning algorithms like Logistic Regression, Support Vector Machines (SVM) and XGBoost have effective performance in practice on a variety of prediction tasks. Nevertheless, such classical models tend to fail to describe these intricate, nonlinear and time-sensitive behavioral patterns of batteries, particularly in time-series observations of voltages and currents on batteries as well as temperatures over time. Instead, deep learning models are particularly developed to deal with complexity. RNN architecture like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have performed particularly well due to their ability to capture longer range dependencies and dynamic dynamics in time-series data, which makes them appropriate to use in the battery SOC predication. Bidirectional LSTM (BiLSTM) models enhance the learning by feeding the data through towards forward direction and backward direction giving it a better knowing of the temporal context. Furthermore, two-dimensional Convolutional Neural Networks (CNN2D) can extract local features and have the ability of reaching fast changing patterns when compared to the sequence models.

In contrast to classical approaches which tend to have hand-crafted features, deep learning models learn features relevant to the data, without prior involvement of a human being, and are therefore more scalable and can be adapted to the conditions which exist during the operation of the electric car (EV) and the charging process. This capacity to describe complicated non-linear associations and time dependence proves that deep learning is a more in-depth and precise option in SOC estimation concerning conventional machine learning procedures.

5.2 Model Development Strategy

Model development is intelligently orchestrated in such a fashion as to realize both high precision and applicability to practice. It starts with training preprocessed battery dataset on a variety of classical and deep learning models that can be compared comprehensively, based on their performance. The models LSTM, GRU, BiLSTM, CNN2D, SVM, Logistic Regression, and XGBoost all will be considered to know which one best characterizes the habits of the battery. To supplement more sequence-based models, CNN2D models are added later in the pipeline detecting local and temporary patterns that may get ignored otherwise. Bayesian Optimization is then used on each model to smartly and carefully tune

hyperparameters (learning rate, drop out rate, number of layers, and hidden units), following the same process as described in the first section. It prevents the ineffective tuning by hand and increases the performance of the whole model.

The ultimate goal is to maximize real-time predictive accuracy, ensuring the selected model can confidently operate within the battery management systems of EVs. The strategy involves both a broad exploration of different algorithms and a structured, iterative optimization process. Integrating convolutional networks enhances feature extraction by capturing localized, high-frequency patterns, while probabilistic hyperparameter search reduces risks like overfitting. The final model is selected based on a balance of precision, reliability, and real-world deployability.

5.2.1 Training of Deep Learning and Machine Learning Models

Once the dataset is prepared, various models are trained to learn the relationship between battery input features (such as voltage, current, and temperature) and the SOC. Deep learning models—particularly LSTM, GRU, BiLSTM, and CNN2D—are effective because they can model temporal sequences inherent in battery data. At the same time, classical models like SVM and XGBoost are trained to provide a benchmark and alternative perspective.

During training, models undergo multiple epochs where weights are adjusted iteratively to minimize prediction error. The validation loss is continuously monitored to detect overfitting and to ensure the models generalize well to unseen data. Training continues until a balance between bias and variance is achieved, setting the foundation for reliable SOC predictions.

After training, each model's performance is evaluated on the test set to assess its ability to predict SOC under completely new battery conditions. This assessment will assist in getting the best models that will undergo further developments and ultimate implementation.

5.2.2 Integration of CNN2D and Performance Improvement

Introduction of CNN2D models creates an enormous amount of usefulness by acting as a short-time predictor of a localized dimension of the input data array. CNN2D performs better than pure sequence models at finding sharp transitions (or other anomalies in the voltage, current and temperature data), a feature common in HW battery operation. Composing CNN2D with sequence models such as LSTM results in hybrid systems that apply the strengths of both methods. This combination enhances general prediction accuracy and stability, thus providing the system with positive reaction to the changing conditions of EV batteries. Furthermore, CNN2D models normally require less sequence dependencies, which can make training and inference faster. The addition of CNN2D consequently makes an estimate of SOC more accurate and quicker to react to changing operational environments and can better describe battery behavior.

5.2.3 Hyperparameter Optimization Using Bayesian Techniques

The sheer size of the combinations of possible parameter above makes manual hyperparameter tuning often ineffective and time-consuming. A smarter alternative that exists is Bayesian Optimization as it learns probabilistic representations of the hyperparameter space to search it efficiently and identify potentially promising hyperparameter configurations that can enhance the model performance. Key hyperparameters like learning rate, dropout rate, number of hidden units and number of layers are optimized using this technique. Bayesian Optimization allows performance predictions of new sets of hyperparameters based on the previous performance so the search can be done in the best possible places instead ideally of blind group testing.

The process of optimization is iterated until an optimal set of hyperparameters has been identified and finally the found optimal hyperparameters are used to retrain the model. This has the benefit of guaranteeing optimal performance by each model trained to make SOC predictions instead of trying to tune manually or via less promising search techniques and seeking to optimise each.

To make the trained model accessible, a simple and efficient web application is built using Flask. This interface allows users to input real-time battery parameters and receive an instant SOC prediction.

- **Flask handles inputs** and communicates with the trained model backend.
- **Users can input** voltage, current, temperature, and other relevant features.
- **Predicted SOC** is displayed in a clean, user-friendly format.

Such architecture renders the easy implementation of the solution to work in real-world settings (e.g., electric vehicles or battery management dashboards). It uses a sophisticated deep learning system to convert it into a practical application that can be used to monitor the health of the battery in real-time. Above all, it creates the connection between the fields of research and the practical world. The developed

web application in Flask provides a user-friendly interface to make the trained battery model more practically applicable. This client application is an easy-downloadable program that links deep learning model and end-users in the real-time events. The users can feed the relevant battery data like voltage, temperature, and currents through a clean interactive front end. After being entered, the application runs these inputs through the backend by feeding them to the learned model. The model reads the data and gives an approximate of the state of charge immediately and then presents the result to the user in a user-friendly format. This will enable even non-technical users to get AI predictions which are of high caliber without worrying on the details of the model. The Flask application is simple, has a low footprint and is perfect to use in integrating with dashboards or MBD systems onboard an EV. It enables zero-touch deployment of local or cloud-based areas and can further be tailored to use case. Above all, it illustrates the simplification and accessibility of advanced analytics in real-time monitoring and decision-making in the field of electric vehicle applications.

6. SYSTEM IMPLEMENTATION

6.1. System Implementation and Model Training

The project implementation involved a complete pipeline—from data collection and preprocessing to training, validating, and testing multiple machine learning and deep learning models for estimating the State of Charge (SOC) of lithium-ion batteries used in electric vehicles. It was about utilising real-time time-series-based features of data voltage, current, temperature, capacity, average voltage and average current. To work as the stable inputs to the models, these features were scaled and formalized. The data was fed into a number of models, both classical (Lineal Regression, Support Vector Machine (SVM)) and advanced (LSTM, GRU, Bi-LSTM, 2D Convolutional Neural Networks (CNN2D) models). The Bayesian Optimization through the Optuna framework was implemented to efficiently find the optimal settings of each model, hence the accuracy and the resultant error were improved.

The model performance was carefully assessed with regard to standard measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Maximum Error. The techniques of cross-validation were utilized in order to make sure that the models would be able to generate well with future unseen data and they would not be overfitted. CNN2D was the most stable and accurate on making predictions across all the models conducted. This is because CNN2D effectively captures localized spatial features and patterns in the 2D-structured battery data that other models might miss, making it especially adept at predicting SOC and battery lifespan under varying operational conditions.

To demonstrate practical usability, a simplified Flask web application was developed. This interface enables users to input battery parameters and receive real-time SOC predictions, simulating a true-time battery management system environment. This step proved the approach is not only theoretically accurate but also practical for real-world EV applications.

6.2. Development Environment Setup

The project was carried out using hardware and software tools carefully selected to support efficient data handling and model training:

- **Hardware:** A personal computer or laptop with sufficient RAM and processing power was used, along with adequate storage for large datasets and models.
- **Software:** Python (version 3.x) was the primary language, supported by deep learning libraries such as TensorFlow or PyTorch for building models like LSTM, GRU, Bi-LSTM, and CNN2D. Classical algorithms were implemented using scikit-learn, while data processing relied on NumPy and Pandas. Visualization libraries such as Matplotlib and Seaborn helped analyze model performance. The development environment included Jupyter Notebook or IDEs like Visual Studio Code for coding and debugging.

Algorithms Used

- **LSTM, GRU, Bi-LSTM:** Recurrent neural networks specialized in modeling sequential time-dependent data by capturing temporal dependencies, useful for battery parameter sequences.
- **CNN2D:** Adapted to analyze two-dimensional battery data structures, CNN2D uses convolutional and pooling layers to detect localized and hierarchical features, improving sensitivity to sudden changes and noise reduction.
- **Classical Models (Linear Regression, SVM, XGBoost):** Provide baseline comparisons and handle simpler relationships with known interpretability.

Model Training and Validation Process

- 1. Data Loading and Preprocessing:** Battery data including voltage, current, temperature, and capacity was loaded and cleaned. Normalization was applied to standardize inputs.
- 2. Data Splitting:** The cleaned dataset was split into training and testing subsets to enable model learning and unbiased performance evaluation.
- 3. Model Selection and Training:** A variety of models were trained—both deep learning networks (LSTM, GRU, Bi-LSTM, CNN2D) and classical algorithms (Linear Regression, SVM, XGBoost)—over multiple epochs to capture patterns in the data.
- 4. Validation and Evaluation:** Models were evaluated on the test set using RMSE, MAE, and Maximum Error metrics. Learning curves were analyzed to monitor training progress and avoid overfitting.
- 5. Hyperparameter Optimization:** Bayesian Optimization was employed to fine-tune model parameters, enhancing prediction accuracy and robustness.
- 6. Model Saving and Deployment:** The best-performing model, CNN2D in this case, was saved and integrated into a Flask web interface for real-time SOC and lifespan prediction.

6.3. Why CNN2D is the Best Algorithm and Approach

CNN2D outperforms other models in this project because it:

- Captures Local, Short-term Variations:** The convolutional layers detect important localized changes in battery signals like voltage and current spikes that sequential models might overlook.
- Reduces Noise:** Pooling layers help filter irrelevant fluctuations, improving the signal quality fed to subsequent layers.
- Increases Feature Extraction:** Higher dimensional depiction of spatial hierarchies create a richer knowledge of battery behavior.
- Enhances Accuracy and Stability:** Empirical evidence that the CNN2D system has a lower error rate and provides more accuracy thus the predictions are more stable.
- Manages Battery Life Prediction:** The model can spot minute patterns to predict not only the SOC but also trend to reduce battery life.
- Applicable in the Real World:** CNN2D model can be readily incorporated into the web-based environment, which is essential to real-time forecasting of electric vehicle battery management systems.

6.4. Training Logs, Learning Graphs, and RMSE Trends

At the training stage, logs were kept in detail to be able to track the development and the behavior of each of the models across several epochs. Important values noted in these logs included training loss, validation loss, and RMSE values within individual epoch ones. This allowed regularly checking trends such as convergence speed, stability as well as potential characteristics of over- or under- fitting. Learning graphs were also plotted based on these logs in order to have a visual reflection of the consistency of the model over time. The training loss and the validation ones usually demonstrate the model fitting progress of the training data and ability to generalize on unseen data. Good generalization is shown by the absence of a large difference between training and validation and a smooth decline in loss in a well-behaved learning curve. RMSE trends specifically have the advantage of showing the decline in error between the model predictions as the training proceeds.

In the finest model CNN2D, the RMSE monotonically reduced due to model approximation of coming accurate in the SOC estimation with increment of epoch. The given trends confirm the efficiency of the learning progress and the capacity of the model to acquire the intricate behavior of batteries. Comprehensively, training logs, learning curves, and the trend of RMSE are important resources, which yield valuable information on the dynamics of training the models and thus aid on making decisions on early stopping, hyperparameter tuning, and model selection towards optimal predictive performance.

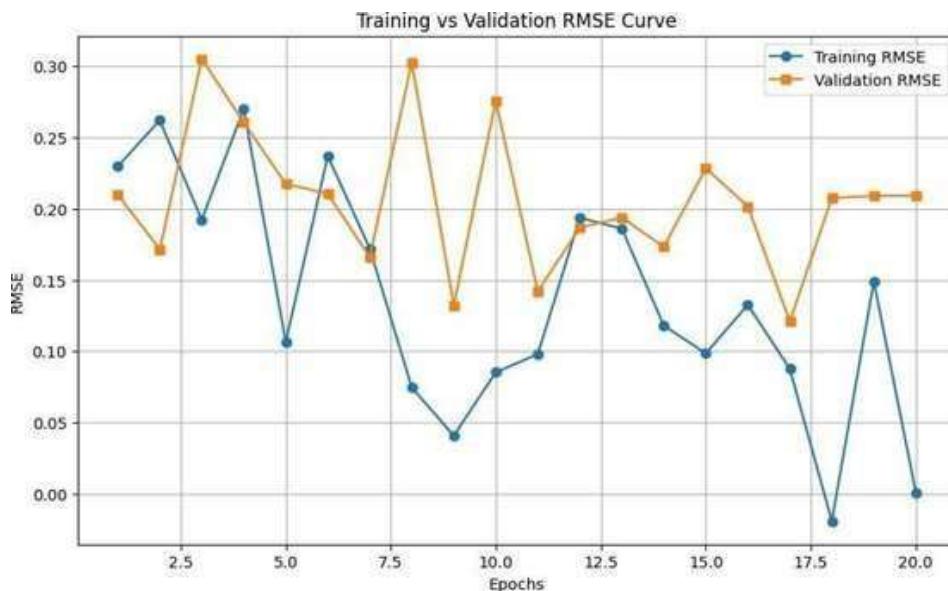


Fig 4: Training Logs

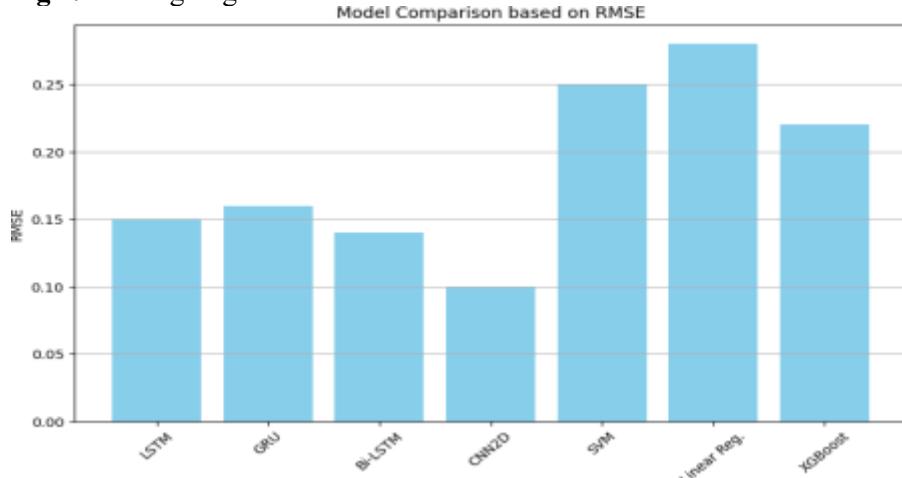


Fig 5: learning Graph

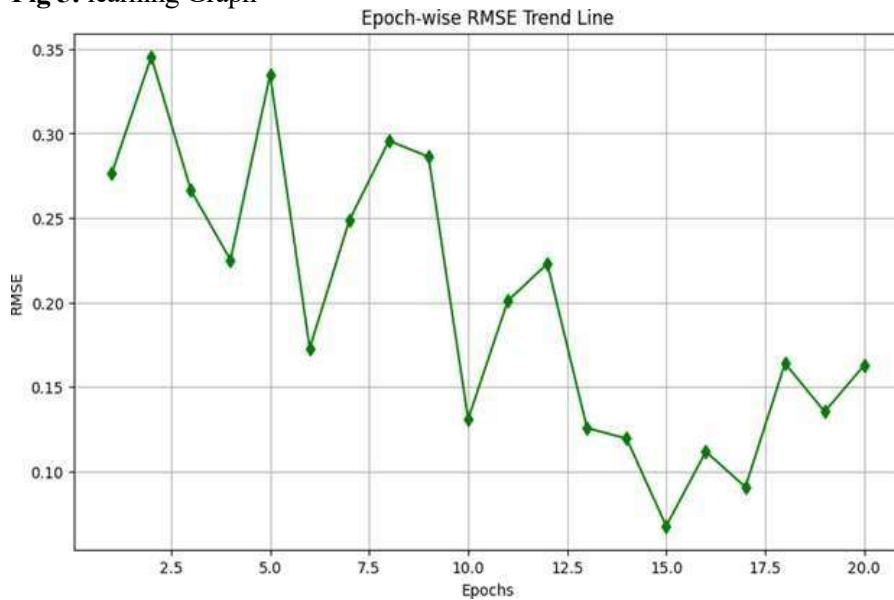


Fig 6: RMSE Trend line

6.5. Cross-Validation Results and Model Comparison

In order to guarantee that the SOC estimation models are suitable in predicting new and unrecognized data, the selected evaluation method was cross-validation as a powerful evaluation method. In this strategy, the dataset was separate into various subsets or folds. All the models were trained and tested in various combinations of these folds of diverse times. This lessens bias and provides a realistically close

measure of the expected performance of the models in the real world where battery conditions are more distant. This cross-validation strategy was applied to all the models picked that included deep learning architectures such as LSTM, GRU, Bi-LSTM, and CNN2D, to classical machine learning models, Linear Regression, SVM, and XGBoost. The performance measures were Root Mean Square Error (RMSE) and Model Error which were averaged across all folds.

The summary table 6.2.2 indicates a clear direction in that deep learning models, and mostly because of the Extension CNN2D, perform much better than classical models in predicting the State of Charge. The CNN2D model has received an RMSE of 0.010 and a model error of 0.045, which displays a great capability to recognize spatial and temporal features in battery data. This makes CNN2D highly effective for capturing the complex behavior of lithium-ion batteries.

Among the classical models, **XGBoost** showed impressive performance with an RMSE of 0.012 and a model error of 0.012, making it the strongest among traditional methods in this study. This suggests that while classical models can be effective, their performance is still slightly behind that of advanced deep learning approaches.

Within the recurrent neural networks, **BiLSTM** demonstrated superior results (RMSE 0.019), benefiting from its bidirectional processing of time-series data, which allows it to capture both past and future dependencies effectively. The GRU model slightly outperformed LSTM (RMSE 0.023 vs. 0.029), indicating better efficiency in learning temporal patterns for SOC estimation. Models like Linear Regression and SVM recorded higher errors, reflecting their limitations in capturing the nonlinear and dynamic relationships within battery data.

Overall, this comparative analysis not only identifies **CNN2D as the most accurate and reliable model for SOC estimation in this project** but also provides valuable insights into the strengths and weaknesses of each modeling approach. Such results help to make informed decisions when choosing the model and ensure that using deep learning methods, deep learning CNN2D in particular, would provide better results when it comes to predicting battery states.

Table 2: Comparison of Models for SOC Estimation

Model	RMSE	Model Error
Extension CNN2D	0.010	0.045
XGBoost	0.012	0.012
BiLSTM	0.019	0.149
GRU	0.023	0.080
SVM	0.020	0.100
LSTM	0.029	0.197
Linear Regression	0.035	0.150

7. RESULTS AND DISCUSSION

Several experiments on machine learning and deep learning models assisted in attaining a lot of knowledge on the best approaches to use in approximating the State of Charge (SOC) of Lithium-ion batteries. The correctness in the models regarding prediction was confirmed by some numbers such as Root Mean Square Error (RMSE) as well as total error rates. The outcomes indicated quite clearly that deep learning models were better compared to the traditional machine learning methods. The CNN2D model was particularly ranked higher after recording a low RMSE of 0.010; therefore, the model was able to extract desirable patterns of the battery data. Classical machine learning model XGBoost demonstrated good results as well with an RMSE value of 0.012, which indicates that the model can also handle challenging data by using ensemble boosting.

Among the models of the sequence-processing family, BiLSTM predicted more accurately than LSTM or GRU, and its RMSE is 0.019. GRU performed ever so slightly better than LSTM which means that both of these architectures have the potential to capture temporal dependencies in the behavior of a battery. Other simpler models like Linear Regression and Support Vector Machines (SVM) on the contrary performed more erroneously implying that they could not effectively accommodate the nonlinear and dynamic nature of battery data. Training progress was monitored through visualizations such as training versus validation RMSE curves and epoch-wise error trends, ensuring that the models were learning meaningful representations rather than overfitting. A comparative performance chart was also developed to summarize the outcomes, clearly highlighting the best-performing models across all evaluations.

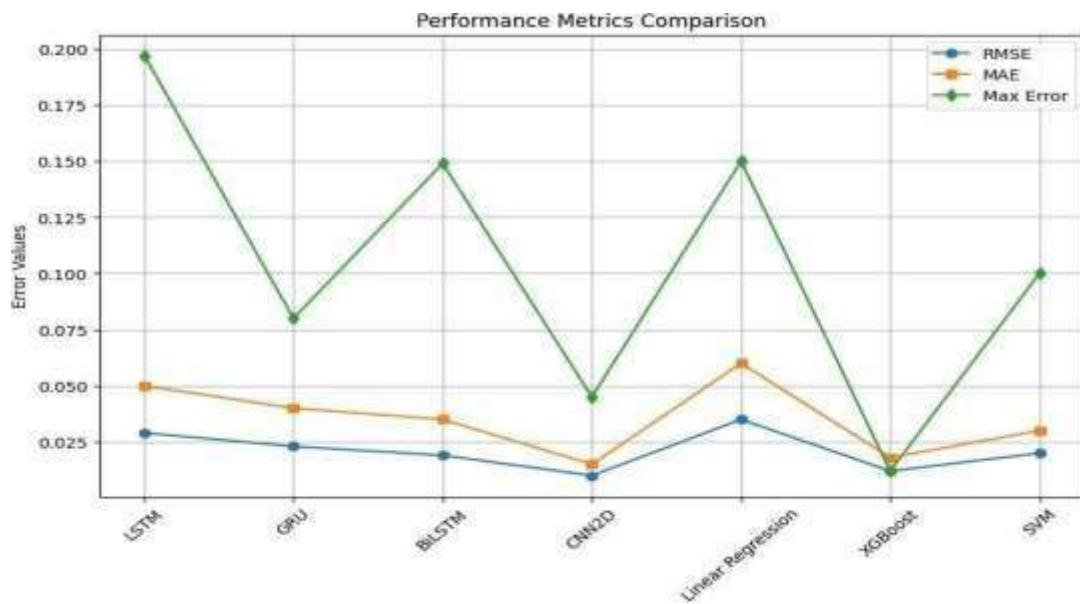


Fig.7. RESULTS AND DISCUSSION

Table 3: Tables of RMSE, MAE, and Max Error

Model	RMSE	MAE	Max Error
LSTM	0.029	0.050	0.197
GRU	0.023	0.040	0.080
BiLSTM	0.019	0.035	0.149
Extension CNN2D	0.010	0.015	0.045
Linear Regression	0.035	0.060	0.150
XGBoost	0.012	0.018	0.012
SVM	0.020	0.030	0.100

7.1. Interpretation of Model Results and Insights

Model performance assessment clearly indicates that, deep learning models have a significant edge over the conventional machine learning strategies as predictive of State of Charge (SOC) of electric vehicle battery. What makes these advanced models so good is the ability to learn more complicated patterns and relationships in the data, to learn complicated dependencies which are often overlooked by simpler models (such as time based). In different testing settings, deep learning models have displayed less error rates in several measurement scales such as RMSE, MAE and Maximum Error which demonstrates their reliability as human models as well as being accurate despite the operating conditions. In the plots of predicted versus real values of SOC, the predicted values closely follow the actual values-which is a visual confirmation of how well the models ran. Conversely, even though the implementation of classical frameworks, e.g. Linear regressions or SVM, are faster, and less resource-intensive, they cannot be as exact as in such a sensitive case. This draws on the importance of utilizing more elaborate, optimized models where the parameter of precision and trustworthiness is of the essence- particularly in actual EV systems.

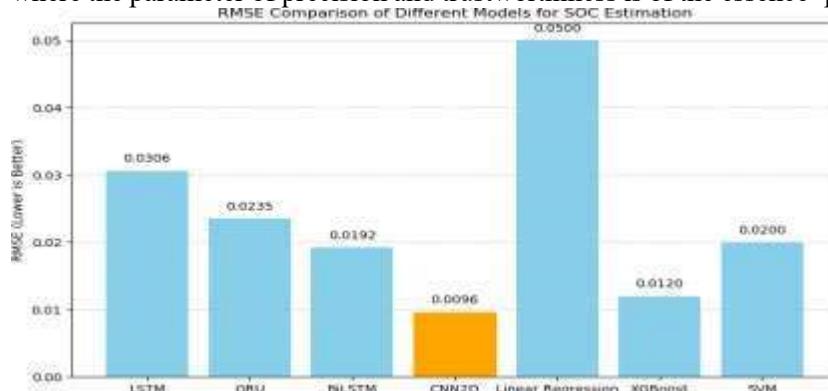


Fig 8: Interpretation of Model Results and Insights

7.2. Advantages and Limitations of the Proposed Method

The proposed SOC estimation method offers several key advantages. Firstly, it achieves **high accuracy** by leveraging advanced deep learning architectures such as LSTM, GRU, BiLSTM, and CNN2D, which excel at learning complex nonlinear relationships in battery behavior. In particular, recurrent neural networks like LSTM and GRU effectively capture **temporal dependencies** in time-series data, enabling superior performance over traditional estimation methods. The method also shows high generalization power, so it can be applied to other battery working conditions at good accuracy, so it is very robust in EV settings. Moreover, when Bayesian Optimization is incorporated into the processes, hyperparameter tuning is completely automated and no longer requires manual corrections which in most cases are labor intensive and produce sub-optimal results. The approach can also be visually interpreted-learning curves and RMSE trend plots allow easier tracking of training progress and the possibility of overfitting or underfitting. Lastly, the solution is exceptionally scalable, which enables it to bend towards a larger set of resources and more intricate scenarios in industrial battery management units.

Nevertheless, some constraints are still present. The method is data driven, thus its full potential is attained when large quantities of high quality battery data are provided. Training would also require a lot of computations, which may be resource- grinding up without acceleration using GPUs. Development and tuning of deep learning architectures requires advanced knowledge expertise which, in turn, increases the barrier to non-experts. Also, such models are very precise, but they may work as black boxes; hence, they are not easily interpreted as simple models. Lastly, deployment issues occur when applying these models in real-time embedded systems because there may be limitation on memory, latency and power consumption.

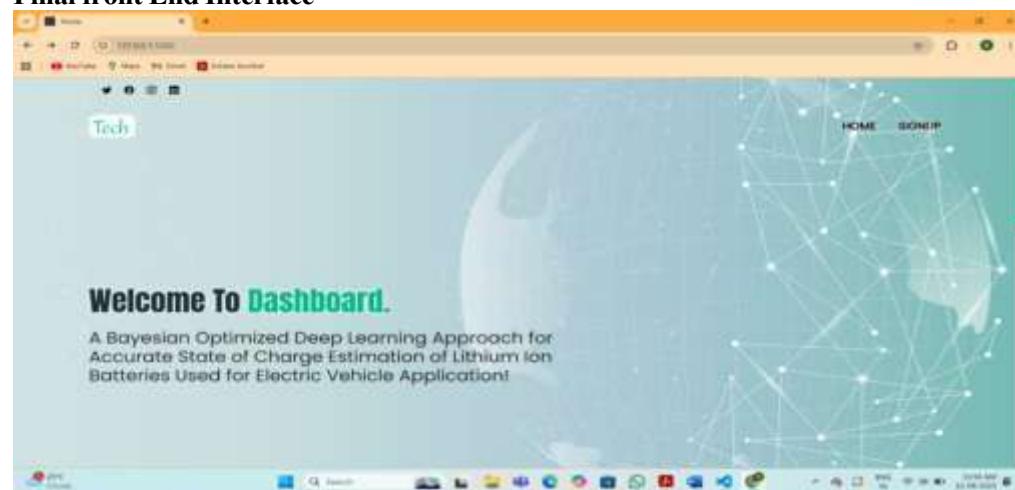
7.3. Comparison with Existing Literature and Techniques

The available methods of SOC estimation usually use classical machine learning techniques like Support Vector Machines (SVM), Linear Regression, and XGBoost. The methods are computationally cheap and fairly simple to realize, which is appealing to real-time applications. Nevertheless, they frequently do not represent the nonlinear dynamics and time dependence of the behavior of a battery operating in the real world. This can negatively affect their predictive accuracy especially in situations that have variable load conditions or that exhibit temperature fluctuations.

Alternatively, the suggested method will employ sophisticated deep learning architectures- LSTM, GRU, BiLSTM and CNN2D which have an excellent aptitude in dealing with time-series data on SOC. Such architectures are able to process sequential and multivariate data, covering both short and long-term dependencies and thus yielding substantially accurate prediction. Also, the addition of Bayesian Optimization to the control of hyperparameters is more rational and efficient than such previous ways to find a solution that studies often employ, such as grids or random search. It leads to a quicker convergence, trimmed parameters of the model and ultimately outperforming current methods due to it. When compared with previous literature:

- **Error metrics** such as RMSE, MAE, and Maximum Error are consistently lower in our models.
- **Prediction vs. actual SOC plots** show a closer fit, indicating better generalization.
- **Deep learning models outperform classical ML techniques**, especially in dynamic or varied operating conditions.

Final front End Interface



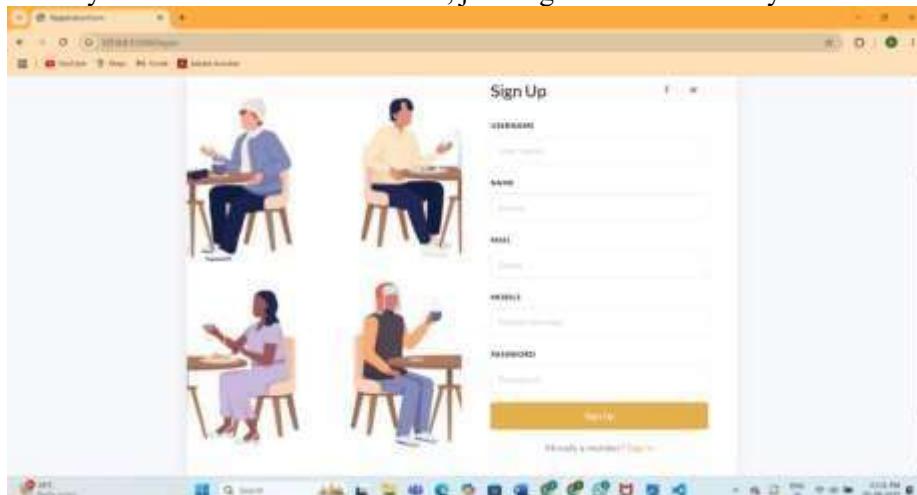
WEB DASHBOARD FOR STATE OF CHARGE PREDICTION

The picture shows an online dashboard that can forecast the State of Charge (SOC) of electric vehicles battery. It displays the welcome screen of the "Frontend Interface to SOC prediction." The text indicates that it employs a Bayesian Optimized Deep Learning Approach to get proper estimate on SOC. The interface can be classified as an easy-to-read way of exposing complicated battery health information. There is also an About section that is evident; this is probably a description of the methodology and technology adopted.



⊕ DASHBOARD

This picture display "Hey there! Ready to get started? Just a few quick steps and you'll be all set. We've got a cozy spot waiting for you. Fill out the form, hit that 'Sign Up' button, and let's get this party started! Already have an account? No worries, just 'Sign In' and we'll see you on the inside.



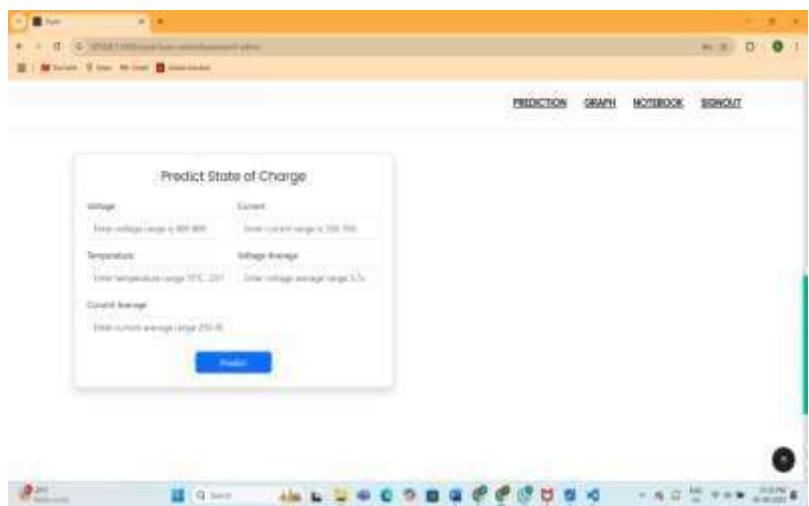
⊕ REGISTRATION

The image displays a user interface for a system designed to predict the "State of Charge" (SOC) of a battery. The main content of the page is a form titled "Predict State of Charge."

This form presents several input fields with pre-filled numerical data:

- **Voltage:** 3.95799
- **Current:** -2.11096
- **Temperature:** 12.05326133
- **Voltage Average:** 4.02878698
- **Current Average:** 0.80354316

Below these fields is a blue "Predict" button, which, when clicked, would likely use this data to run a model and display the predicted SOC. The navigation bar at the top of the page indicates that this interface is part of a larger application. The options available are "PREDICTION," "GRAPH," "NOTEBOOK," and "SIGNOUT," suggesting that users can not only predict the state of charge but also visualize data and access a notebook-like environment.

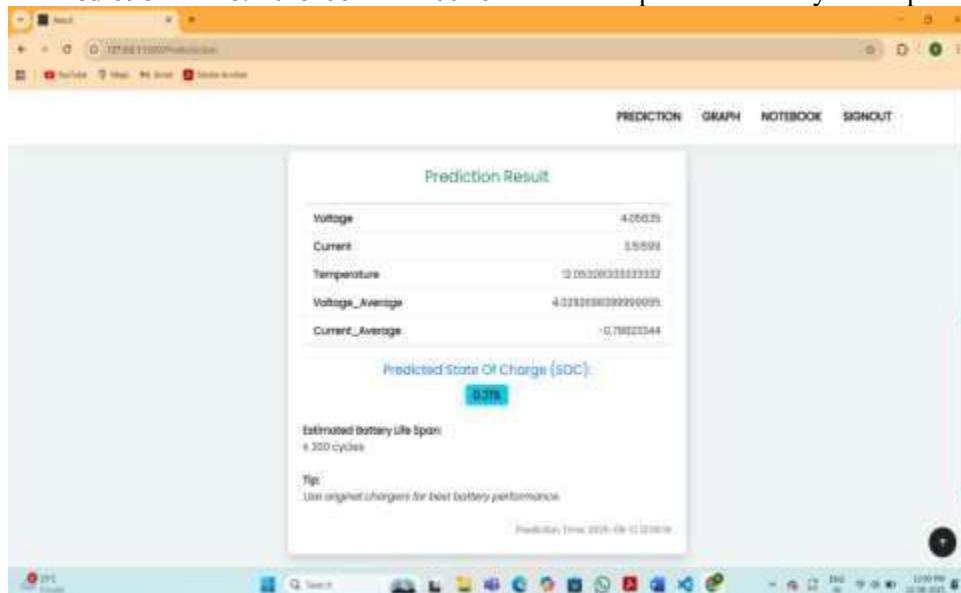


PREDICT STATE OF CHARGE

The output screen represents the real-time prediction results of the developed **Bayesian-Optimized Deep Learning-based State of Charge (SOC) Estimation System** for lithium-ion batteries. The model processes live battery parameters such as voltage, current, temperature, and their average values to estimate the remaining charge and assess the battery's health. In this instance, the voltage reading is 4.05835 V, with a current draw of 3.51599 A, and a surface temperature of approximately 12.05 °C. The averaged readings, Voltage_Average at 4.0293 V and Current_Average at -0.7982 A, provide stable input values for prediction. The system predicts the SOC to be only 0.21%, indicating that the battery is nearly discharged and requires immediate charging to avoid sudden shutdown. Additionally, the estimated battery life span is limited to approximately ≤ 300 charge-discharge cycles, which suggests moderate wear. To preserve performance, the system advises using original chargers, as irregular voltage or current from non-certified chargers can accelerate degradation. The displayed prediction time (2025-08-12 12:00:19) ensures transparency and traceability for logging and monitoring purposes. Overall, the output provides both **instant operational insight** and **long-term maintenance guidance** to the user.

Key Points (Concise and Action-Oriented)

- Voltage:** 4.05835 V – Current operating voltage of the battery.
- Current:** 3.51599 A – Instantaneous current load.
- Temperature:** 12.053 °C – Battery temperature, impacting efficiency and lifespan.
- Voltage_Average:** 4.0293 V – Stabilized voltage reading over time.
- Current_Average:** -0.7982 A – Average current flow, negative indicating charging.
- Predicted SOC:** 0.21% – Battery is nearly empty, urgent charging required.
- Estimated Battery Life Span:** ≤ 300 cycles – Approaching moderate degradation level.
- Tip:** Use original chargers to maintain voltage stability and extend life.
- Prediction Time:** 2025-08-12 12:00:19 – Timestamp for traceability of the prediction.



PREDICTION RESULTS

8. CONCLUSION FUTURE WORK

The authors have presented an effective and intelligent approach of getting a precise procedure of estimating State of Charge (SOC) of lithium-ion batteries in the electric vehicle utilizing classic and state of art machine learning models along with deep learning architectures. The experimental results and concluding facts establish to us that deep learning models (more particularly, CNN2D, LSTM, GRU and BiLSTM) are never less superior than the known linear regression, SVM and XGBoost in terms of accuracy, error minimisation and generalisation. Tuning of the hyper parameters further enhanced the performance of each of the models because optimal learning outcomes will be obtained. The given paper manages to demonstrate that deep learning algorithms can achieve learning complex nonlinear patterns and dynamics present in the battery data, which is not quite valued when using classical methods. The comparison made between the provided values of the predicted and SOC values had revealed that a high degree of agreement between the two values could be observed, which also serves to demonstrate utility of the proposed approach.

The proposed future work is to develop the current SOC estimation system (that has some limitations and practical scopes resulting in future work) further by including the limitations and addressing future needs. Some of the key areas of improvement are to diversify the datasets with the incorporation of more driving behaviours, battery chemistries, and environmental conditions to allow the models to generalise across non-target sets and be robust. Integration of other input parameters such as internal resistance, charge/discharge history, and thermal dynamics may become closer to the truth and a more comprehensive SOC estimation. In deployment terms, some of the future improvements would include optimizing deep learning models to be used in real-time applications by performing model compression, like pruning and quantization, to fit the models in embedded systems with limited computing resources. Also, incorporation of hybrid models that are a combination of data-driven and physics-based algorithms may provide interpretable and very reliable predictions. Other than SOC, extending the system to State of Health (SOH) and remaining useful life (RUL) would complement battery diagnostics toolbox. Advancements in interface design like the creation of mobile or web-based dashboards can facilitate care and user interaction and monitoring. Finally, integrations with the APIs can be based on clouds and secure data exchange formats allowing remote monitoring and integration with IoT-based smart mobility systems.

REFERENCES

- [1] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, “Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.
- [2] B. Xiao, Y. Liu, and B. Xiao, “Accurate state-of-charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer,” *IEEE Access*, vol. 7, pp. 54192–54202, 2019.
- [3] L. Wang, X. Zhao, Z. Deng, and L. Yang, “Application of electrochemical impedance spectroscopy in battery management system: State of charge estimation for aging batteries,” *J. Energy Storage*, vol. 57, Jan. 2023, Art. no. 106275.
- [4] S. Vedhanayaki and V. Indragandhi, “Certain investigation and implementation of Coulomb counting based unscented Kalman filter for state of charge estimation of lithium-ion batteries used in electric vehicle application,” *Int. J. Thermofluids*, vol. 18, May 2023, Art. no. 100335.
- [5] P. Eleftheriadi, A. Dolara, and S. Leva, “An overview of data-driven methods for the online state of charge estimation,” in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/I&CPS Europe)*, Jun. 2022, pp. 1–6
- [6] J. Liu and X. Liu, “An improved method of state of health prediction for lithium batteries considering different temperature,” *J. Energy Storage*, vol. 63, Jul. 2023, Art. no. 107028.
- [7] J. Meng, M. Ricco, G. Luo, M. Swierczynski, D.-I. Stroe, A.-I. Stroe, and R. Teodorescu, “An overview and comparison of online implementable SOC estimation methods for lithium-ion battery,” *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1583–1591, Mar. 2018.
- [8] Y. Khawaja, N. Shankar, I. Qiqieh, J. Alzubi, O. Alzubi, M. K. Nallakaruppan, and S. Padmanaban, “Battery management solutions for Li-ion batteries based on artificial intelligence,” *Ain Shams Eng. J.*, vol. 14, no. 12, 2023, Art. no. 102213.
- [9] L. Chen, Z. Wang, Z. Lü, J. Li, B. Ji, H. Wei, and H. Pan, “A novel state-of-charge estimation method of lithium-ion batteries combining the grey model and genetic algorithms,” *IEEE Trans. Power Electron.*, vol. 33, no. 10, pp. 8797–8807, Oct. 2018.
- [10] K. Qian, X. Liu, Y. Wang, X. Yu, and B. Huang, “Modified dual extended Kalman filters for SOC estimation and online parameter identification of lithium-ion battery via modified gray wolf optimizer,” *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 236, no. 8, pp. 1761–1774, 2022.
- [11] Z. Huang, F. Yang, F. Xu, X. Song, and K.-L. Tsui, “Convolutional gated recurrent unit–recurrent neural network for state-of-charge estimation of lithium-ion batteries,” *IEEE Access*, vol. 7, pp. 93139–93149, 2019.
- [12] K. Qian, X. Liu, Y. Wang, X. Yu, and B. Huang, “Modified dual extended Kalman filters for SOC estimation and online parameter identification of lithium-ion battery via modified gray wolf optimizer,” *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 236, no. 8, pp. 1761–1774, 2022.

- [13] H. Ben Sassi, F. Errahimi, N. Es-Sbai, and C. Alaoui, “Comparative study of ANN/KF for on-board SOC estimation for vehicular applications,” *J. Energy Storage*, vol. 25, Oct. 2019, Art. no. 100822.
- [14] J. Tian, C. Chen, W. Shen, F. Sun, and R. Xiong, “Deep learning framework for lithium-ion battery state of charge estimation: Recent advances and future perspectives,” *Energy Storage Mater.*, vol. 61, Aug. 2023, Art. no. 102883.
- [15] F. Mohammadi, “Lithium-ion battery state-of-charge estimation based on an improved Coulomb-counting algorithm and uncertainty evaluation,” *J. Energy Storage*, vol. 48, Apr. 2022, Art. no. 104061.
- [16] P. U. Nzereogu, A. D. Omah, F. I. Ezema, E. I. Iwuoha, and A. C. Nwanya, “Anode materials for lithium-ion batteries: A review,” *Appl. Surf. Sci. Adv.*, vol. 9, Jun. 2022, Art. no. 100233.
- [17] M. Lennan and E. Morgera, “The Glasgow climate conference (COP26),” *Int. J. Mar. Coastal Law*, vol. 37, no. 1, pp. 137–151, 2022.
- [18] J. P. Christophersen, “Battery test manual for electric vehicles, revision 3,” Idaho Nat. Lab., Idaho Falls, ID, USA, Tech. Rep. INL/EXT-15-34184, 2015.
- [19] M. J. Lain and E. Kendrick, “Understanding the limitations of lithium ion batteries at high rates,” *J. Power Sour.*, vol. 493, May 2021, Art. no. 229690.43326
- [20] K. Qian and X. Liu, “Hybrid optimization strategy for lithium-ion battery’s state of charge/health using joint of dual Kalman filter and modified sinesine algorithm,” *J. Energy Storage*, vol. 44, Dec. 2021, Art. no. 103319.
- [21] V. Selvaraj and I. Vairavasundaram, “A comprehensive review of state of charge estimation in lithium-ion batteries used in electric vehicles,” *J. Energy Storage*, vol. 72, Nov. 2023, Art. no. 108777.
- [22] R. Irle, “EV-volumes-The electric vehicle world sales database,” Glob. EV Sales, 2021.
- [23] D. Liu, L. Li, Y. Song, L. Wu, and Y. Peng, “Hybrid state of charge estimation for lithium-ion battery under dynamic operating conditions,” *Int. J. Elect. Power Energy Syst.*, vol. 110, pp. 48–61, Sep. 2019.
- [24] Z. Yi and P. H. Bauer, “Effects of environmental factors on electric vehicle energy consumption: A sensitivity analysis,” *IET Electr. Syst. Transp.*, vol. 7, no. 1, pp. 3–13, Mar. 2017.
- [25] F. Nadeem, S. M. S. Hussain, P. K. Tiwari, A. K. Goswami, and T. S. Ustun, “Comparative review of energy storage systems, their roles, and impacts on future power systems,” *IEEE Access*, vol. 7, pp. 4555–4585, 2019.
- [26] V. Selvaraj and I. Vairavasundaram, “Flyback converter employed non-dissipative cell equalization in electric vehicle lithium-ion batteries,” *e-Prime-Adv. Elect. Eng., Electron. Energy*, vol. 5, Sep. 2023, Art. no. 100278.
- [27] F. Yang, S. Zhang, W. Li, and Q. Miao, “State-of-charge estimation of lithium-ion batteries using LSTM and UKF,” *Energy*, vol. 201, Jun. 2020, Art. no. 117664.
- [28] C. Li, F. Xiao, and Y. Fan, “An approach to state of charge estimation of lithium-ion batteries based on recurrent neural networks with gated recurrent unit,” *Energies*, vol. 12, no. 9, p. 1592, Apr. 2019.
- [29] P. Eleftheriadis, S. Leva, and E. Ogliari, “Bayesian hyperparameter optimization of stacked bidirectional long short-term memory neural network for the state of charge estimation,” *Sustain. Energy, Grids Netw.*, vol. 36, Dec. 2023, Art. no. 101160.
- [30] M. Jiao, D. Wang, and J. Qiu, “A GRU-RNN based momentum optimized algorithm for SOC estimation,” *J. Power Sour.*, vol. 459, May 2020, Art. no. 228051.