

Review

Not peer-reviewed version

---

# Machine Learning-Based Regression Models for State of Charge Estimation in Hybrid Electric Vehicles: A Review

---

[Arash Mousaei](#)<sup>\*</sup> and Yahya Naderi

Posted Date: 26 December 2023

doi: 10.20944/preprints202312.1938.v1

Keywords: Hybrid Electric Vehicles (HEVs); State of Charge (SOC); MACHINE LEARNING; Support Vector Regression (SVR); Neural Network Regression (NNR)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Machine Learning-Based Regression Models for State of Charge Estimation in Hybrid Electric Vehicles: A Review

Arash Mousaei <sup>1,\*</sup> and Yahya Naderi <sup>2</sup>

<sup>1</sup> Dept. Electrical and Computer Engineering University of Tabriz Tabriz, Iran

<sup>2</sup> Dept. Electronic and Electrical Engineering University of Strathclyde Glasgow, UK; y.naderi@strath.ac.uk

\* Correspondence: a.mousaei@tabrizu.ac.ir

**Abstract:** This article explores the crucial task of estimating State of Charge (SOC) in Hybrid Electric Vehicles (HEVs) and examines the applicability of various regression models for this purpose. We delve into the strengths and limitations of Linear Regression, Support Vector Regression (SVR), and Neural Network Regression (NNR) in the context of SOC estimation. Linear regression provides a simple and interpretable baseline, SVR extends this to nonlinear relationships, while NNR emerges as a powerful tool with adaptive capabilities. The choice of model depends on factors such as data characteristics, interpretability, and computational resources. As the field evolves, the article advocates for a nuanced approach, possibly incorporating hybrid models, to achieve robust and accurate SOC estimation, contributing to the ongoing enhancement of HEV efficiency and sustainability.

**Keywords:** Hybrid Electric Vehicles (HEVs); State of Charge (SOC); MACHINE LEARNING; Support Vector Regression (SVR); Neural Network Regression (NNR)

---

## I. Introduction

In the rapidly evolving landscape of transportation, the development of energy-efficient and sustainable mobility solutions has become imperative. Hybrid Electric Vehicles (HEVs) stand at the forefront of this paradigm shift, offering a harmonious blend of conventional and electric powertrains. Central to the optimal functioning of HEVs is the precise estimation of the State of Charge (SOC) of their high-voltage battery systems [1]. SOC, representing the remaining energy capacity in the battery, plays a pivotal role in orchestrating the delicate balance between electric and internal combustion propulsion systems within these vehicles [2]. The urgency for accurate SOC estimation arises from the multifaceted challenges faced by traditional methods [3]. Voltage-based methods, ampere-hour counting, and open-circuit voltage methods, while providing reasonable estimates in controlled environments, falter when exposed to the complexities of real-world driving conditions [4]. As HEVs traverse dynamic terrains, experience varying environmental conditions, and utilize batteries with diverse chemistries, the limitations of traditional SOC estimation methods become increasingly apparent [5].

Traditional methods often struggle to accurately estimate SOC when confronted with the nonlinear discharge and charge patterns inherent in HEV battery operation [6]. This limitation can lead to suboptimal energy utilization and compromise the overall efficiency of the vehicle. The sensitivity of SOC estimation to environmental factors, especially temperature, poses a considerable challenge [7]. Traditional methods may inadequately compensate for these variations, resulting in estimation errors and diminished reliability in real-world driving scenarios. The diverse landscape of battery chemistries employed in HEVs requires adaptable SOC estimation methods [8]. Traditional approaches may lack the flexibility needed to cater to the unique characteristics of different battery types, limiting their applicability [9]. HEVs frequently operate under dynamic conditions, including

regenerative braking, rapid acceleration, and load variations [10]. Traditional methods struggle to capture the rapid fluctuations in SOC under these scenarios, impacting the accuracy of estimation. Addressing these challenges necessitates a paradigm shift in SOC estimation methodologies. The integration of machine learning-based regression models emerges as a promising solution, offering a more nuanced and adaptable approach to the intricacies of SOC estimation in HEVs [11].

Machine learning, a subfield of artificial intelligence, empowers systems to learn and adapt from data, providing a dynamic framework for SOC estimation [12]. Within the realm of machine learning, regression models stand out as potent tools for predicting continuous variables, making them well-suited for the task of estimating SOC [13]. Machine learning-based regression models exhibit a remarkable ability to adapt to diverse battery chemistries, dynamic operating conditions, and environmental variability [14]. This adaptability addresses the shortcomings of traditional methods and enhances the accuracy and robustness of SOC estimation [15]. The nonlinear relationship between battery parameters and SOC is more effectively captured by regression models [16]. Traditional linear methods struggle to represent the intricate dependencies within the data, limiting their ability to provide accurate estimates [17]. Some machine learning models offer the potential for real-time SOC estimation, a critical feature for optimizing energy management and ensuring responsive control in HEV operations [18].

The primary objective of this comprehensive review paper is to delve into the nuances of machine learning-based regression models for SOC estimation in HEVs. By providing an in-depth exploration of methodologies, key findings, and implications, this paper aims to contribute to the growing body of knowledge surrounding the convergence of machine learning and sustainable transportation.

The subsequent sections of this paper are meticulously structured to provide a comprehensive understanding of the application of machine learning-based regression models for SOC estimation in HEVs. Section 2 offers a detailed background on HEVs, emphasizing the pivotal role of SOC in their operation. Section 3 delves into the realm of machine learning for SOC estimation, elucidating the significance of regression models and their application in the context of HEVs. In section 4, we will have results. In summary, this paper endeavors to serve as a beacon in the exploration of machine learning-based regression models for SOC estimation in HEVs, offering insights and perspectives that contribute to the ongoing discourse on sustainable transportation and energy-efficient mobility solutions.

## II. Background of soc in hevs

HEVs represent a pivotal development in the evolution of the automotive industry, designed to bridge the gap between traditional internal combustion engine vehicles and fully electric vehicles [3]. These innovative vehicles incorporate both an internal combustion engine and an electric motor, offering improved fuel efficiency, reduced emissions, and increased energy efficiency compared to their conventional counterparts. Fundamental to the effective operation of HEVs is the precise monitoring and estimation of the SOC of the vehicle's energy storage system, typically a high-voltage battery pack [8].

### A. Definition and Significance of SOC

SOC is a critical parameter that indicates the amount of energy stored in the battery of an HEV at any given moment (equation (1)). It is often expressed as a percentage, where SOC signifies a completely discharged battery (0%) and a fully charged battery (100%). SOC estimation is essential because it enables the vehicle's control systems to make informed decisions about the use of electric power, optimizing fuel efficiency and extending the life of the battery [11].

$$SOC (\%) = \frac{Q_{used}}{Q_{total}} \times 100 \quad (1)$$

Where, SOC% is the State of Charge as a percentage,  $Q_{used}$  is the amount of charge used, and  $Q_{total}$  is the total charge capacity of the battery.

Precise SOC estimation is crucial for several reasons:

- (1) **Energy Management:** HEVs operate by seamlessly switching between the internal combustion engine and the electric motor, depending on driving conditions and power demands. Accurate SOC estimation is instrumental in determining when to engage the electric motor or the internal combustion engine to ensure optimal energy management [3].
- (2) **Battery Health and Longevity:** Inaccurate SOC estimation can lead to inadequate charging or discharging of the battery, potentially causing overcharging or deep discharging. Both of these scenarios can compromise battery health and significantly reduce its lifespan. Proper SOC management helps maintain battery health and prolong its operational life [5].
- (3) **Fuel Efficiency:** Precise SOC estimation is crucial for maximizing fuel efficiency in HEVs. It ensures that electric power is used effectively during low-load conditions, reducing the reliance on the internal combustion engine and minimizing fuel consumption [6].
- (4) **Emissions Reduction:** HEVs are renowned for their reduced emissions compared to traditional vehicles. Accurate SOC estimation plays a pivotal role in enabling the vehicle to operate in electric-only mode when possible, further reducing emissions and promoting environmental sustainability [9].

### *B. Challenges in SOC Estimation*

Despite the significance of SOC estimation, traditional methods have faced several challenges, hindering their accuracy and adaptability to the diverse operating conditions that HEVs encounter. Traditional methods for SOC estimation include voltage-based methods, ampere-hour counting, and open-circuit voltage methods. While these methods can provide reasonable estimates under specific conditions, they struggle to overcome the following challenges:

- (1) **Inaccuracies:** Traditional methods often result in inaccurate SOC estimates, particularly when the battery's discharge and charge patterns are nonlinear. These inaccuracies can lead to suboptimal vehicle performance, reduced fuel efficiency, and decreased battery utilization [10].
- (2) **Environmental Variability:** SOC estimation is sensitive to environmental factors, especially temperature. Variations in temperature can significantly affect battery performance and, consequently, SOC estimation. Traditional methods may not adequately account for these variations, resulting in estimation errors [14].
- (3) **Complex Battery Chemistry:** HEV batteries employ various chemistries, each with its unique characteristics [17]. Traditional methods may not adapt well to the specific behaviors of different battery types, limiting their versatility.
- (4) **Dynamic Operating Conditions:** HEVs frequently operate in diverse and dynamic conditions, including regenerative braking, fast acceleration, and variations in load. Traditional methods may not accurately capture these dynamic changes in SOC, potentially leading to inaccurate estimations [19].

In response to these challenges, recent advancements have turned toward machine learning-based regression models to address the intricacies of SOC estimation in HEVs. These models offer adaptability, accuracy, and robustness, which are essential for navigating the diverse challenges that arise in the operation of HEVs. This paper will delve further into the application of machine learning-based regression models for SOC estimation in HEVs, examining their methodologies, key findings, and implications in depth.

### **III. Machine Learning for SOC Estimation in HEVs**

In the pursuit of refining SOC estimation in HEVs, the integration of machine learning represents a paradigm shift that transcends the limitations of traditional methods. This section delves into the realm of machine learning, elucidating the significance of regression models and their application in the context of HEVs.

### A. The Essence of Machine Learning in SOC Estimation

Machine learning, a subfield of artificial intelligence, empowers systems to learn from data and make informed decisions without explicit programming [20]. In the context of SOC estimation for HEVs, machine learning techniques offer a dynamic and adaptable approach, capable of learning intricate patterns and relationships within the data [21]. Machine learning's capacity for adaptation to diverse conditions, nonlinear relationship modeling, and real-time processing aligns seamlessly with the intricacies of SOC estimation in HEVs [22]. Unlike traditional methods that rely on predefined algorithms, machine learning models can evolve and improve their performance over time, making them well-suited for the dynamic nature of HEV operations [23].

### B. The Significance of Regression Models in SOC Estimation

Regression models within the domain of machine learning play a pivotal role in predicting continuous variables, making them particularly apt for estimating SOC. These models establish a mathematical relationship between input features and the target variable, allowing for the precise prediction of SOC based on the given data [24].

#### 1) Linear Regression:

Linear regression, a fundamental type of regression model, assumes a linear relationship between the input features and the target variable. While simplistic, it provides a baseline for understanding the general trend and can be a valuable starting point for SOC estimation (Figure 1 and equation (2)) [25].

$$SOC = \beta_0 + \beta_1 \times Feature_1 + \beta_2 \times Feature_2 + \dots + \beta_n \times Feature_n + \varepsilon \quad (2)$$

Where,  $\beta_0, \beta_1, \beta_n$  are the coefficients,  $Feature_1, Feature_2, Feature_n$  are the input features, and  $\varepsilon$  is the error term.

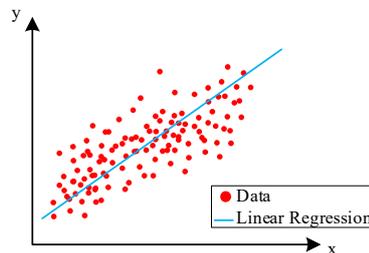


Figure 1. Linear Regression.

The use of linear regression for estimating SOC in HEVs comes with several advantages. Linear regression provides a clear and interpretable relationship between the independent variables (e.g., current, voltage, temperature) and the SOC. The coefficients in the regression equation indicate the magnitude and direction of the influence of each variable, offering valuable insights into the charging and discharging dynamics of the battery [26]. The linear regression model is relatively simple and computationally efficient compared to more complex modeling techniques. This simplicity facilitates easier implementation, understanding, and integration into real-time control systems of HEVs. Linear regression models can be easily updated and refined as new data becomes available [27]. This adaptability ensures that the model remains accurate and relevant over time, accommodating changes in battery behavior or system characteristics [28]. Linear regression effectively utilizes historical data to establish relationships between independent variables and SOC. This utilization of past information allows for the creation of a robust and reliable model, enhancing its predictive capabilities [29]. The simplicity of linear regression makes it computationally efficient, requiring fewer computational resources compared to more complex modeling techniques. This efficiency is particularly advantageous for on-board systems in HEVs, where computational resources may be limited [30]. Linear regression can be applied not only to estimate SOC but also to investigate the impact of different factors on battery performance. This versatility makes it a valuable tool for

researchers and engineers seeking a comprehensive understanding of the interactions within the hybrid system [31]. The simplicity and efficiency of linear regression contribute to its cost-effectiveness, both in terms of model development and implementation. This makes it an attractive option for applications where cost considerations are significant [32].

While linear regression offers several advantages for estimating SOC in HEVs, it's essential to consider its limitations. Linear regression assumes a linear relationship between the independent variables and SOC [33]. If the true relationship is nonlinear, the model may not accurately capture the complexities of the system, leading to potential inaccuracies in predictions [34]. Linear regression is sensitive to outliers in the data. Unusual or extreme data points can disproportionately influence the regression coefficients, impacting the overall model performance [35]. The simplicity of linear regression, which is often an advantage, can be a limitation when dealing with highly complex systems. If the relationships between variables are intricate and nonlinear, a more sophisticated modeling technique may be required for accurate predictions [36]. Achieving the right balance in model complexity is crucial. Overfitting (capturing noise in the training data) or underfitting (oversimplifying the model) can both lead to poor generalization performance on new, unseen data [37]. Linear regression assumes that the residuals (the differences between predicted and actual values) are independent. If there is autocorrelation or dependence among residuals, it can affect the model's accuracy and reliability. Linear regression is designed for numeric data and may not be suitable for categorical variables without appropriate transformations [38]. This limitation may be relevant when dealing with non-numeric aspects of battery behavior. Linear regression assumes that the residuals are normally distributed. If the residuals deviate significantly from a normal distribution, it may affect the validity of statistical inferences [39]. The performance of a linear regression model heavily relies on the quality and representativeness of the training data. If the data is biased or not reflective of the full range of operating conditions, the model's predictions may lack accuracy. Linear regression models assume a static relationship between variables, which may not capture changes in battery behavior over time or due to aging effects. Dynamic and evolving systems may require more sophisticated modeling techniques [40]. All of these advantages and disadvantages were summarized in the Table I.

**TABLE I.** THE ADVANTAGES AND DISADVANTAGES OF LINEAR REGRESSION.

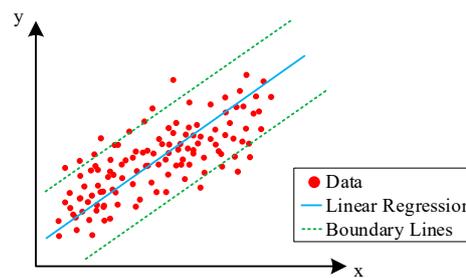
Ref.	Advantages	Disadvantages
	Interpretability	Linearity Assumption
[26]	Simplicity	Sensitivity to Outliers
	Real-time Predictions	Limited Complexity
	Adaptability	Overfitting/Underfitting
[27]	Data Utilization	Assumption of Independence
	Resource Efficiency	Limited to Numeric Data
[28]	Versatility	Limited in Handling Non-Gaussian Residuals
	Cost-Effectiveness	Data Quality Dependency
[29]	Ease of Implementation	Static Nature
[30]	Versatile in Variable Types	Multicollinearity
[31]	Model Transparency	Limited for Time-Series Data
[32]	Assumes Homoscedasticity	Dependence on Training Data
[33]	Efficient for Large Datasets	Assumption of Normality
[34]	Facilitates Hypothesis Testing	Limited for Complex Systems
[35]	Ease of Model Interpretation	Limited Feature Engineering
[36]	Useful for Exploratory Analysis	Limited in Handling Missing Data
[37]	Robust to Irrelevant Features	Difficulty with Non-Continuous Variables

## 2) Support Vector Regression (SVR):

Support Vector Regression (SVR) is a non-linear regression model that leverages support vector machines to capture complex relationships in the data (Figure 2 and equation 3). The kernel function allows SVR to map the input features into a higher-dimensional space, facilitating the modeling of intricate SOC patterns [41].

$$SOC = \sum_{i=1}^n (\alpha_i \times K(X_i, X) + b) + \varepsilon \quad (3)$$

That,  $\alpha_i$  are the support vector coefficients,  $K(X_i, X)$  is the kernel function, and  $b$  is the bias term.



**Figure 2.** Support Vector Regression.

SVR is a powerful machine learning technique that offers several advantages for estimating SOC in HEVs. SVR can capture nonlinear relationships between independent variables (such as current, voltage, and temperature) and SOC. This is particularly beneficial when the true relationships in the data are complex and cannot be adequately represented by linear models [27]. SVR is effective in high-dimensional spaces, making it suitable for applications where the number of independent variables is large. This is advantageous in the context of HEVs, where multiple factors can influence the battery's state [30]. SVR uses the kernel trick, allowing it to implicitly map input data into higher-dimensional feature spaces. This can enhance the model's ability to capture intricate patterns in the data, providing a more accurate representation of the underlying relationships [33]. SVR is less sensitive to outliers in the data compared to traditional linear regression methods. The robustness to outliers is particularly valuable in real-world scenarios where data may contain noise or anomalies [35]. SVR aims to find a hyperplane that best fits the data while minimizing the error. This global optimization approach can lead to more accurate and generalized models, reducing the risk of overfitting to specific data points [36]. SVR allows for the use of different kernel functions, such as linear, polynomial, and Radial Basis Function (RBF) kernels. This flexibility enables the model to adapt to various data patterns and distributions [37]. SVR includes tuning parameters, such as the regularization parameter ( $\alpha_i$ ) and kernel parameters. Proper tuning allows for optimization of the model's performance, making it adaptable to different datasets [38]. SVR uses a subset of training data points, known as support vectors, to define the regression function. This memory-efficient approach makes SVR suitable for large datasets without compromising performance [39]. SVR can perform well even when the number of training samples is relatively small. This is advantageous in situations where collecting extensive training data may be challenging or expensive [40]. SVR often provides high prediction accuracy, especially in situations where the relationships between variables are complex and nonlinear. This accuracy is crucial for reliable SOC estimation in HEVs. SVR tends to generalize well to unseen data, making it robust in scenarios where the model needs to perform accurately on new, previously unseen data points [41].

While SVR has several advantages, it also comes with certain disadvantages that need to be considered in the context of estimating SOC in HEVs. SVR can be computationally intensive, especially when dealing with large datasets or complex kernel functions. Training the model may require significant time and computational resources, making it less practical for real-time applications in certain cases [27]. The effectiveness of SVR is highly dependent on proper parameter

tuning, including the choice of the kernel and tuning parameters. Selecting the appropriate parameters can be challenging, and an improperly tuned model may lead to overfitting or underfitting [29]. SVR models, particularly with complex kernel functions, can be challenging to interpret. Understanding the impact of individual features on the prediction may not be as straightforward as in simpler models like linear regression [30]. While SVR is memory-efficient due to its reliance on support vectors, the storage requirements can still be substantial, especially when dealing with large datasets. This can limit its applicability in resource-constrained environments [32]. SVR can be sensitive to noise in the training data, especially when the dataset contains outliers. Outliers may become support vectors and disproportionately influence the model, potentially leading to suboptimal performance [33]. The inner workings of SVR, especially with complex kernels, can be considered a "black box." This lack of transparency may be a drawback in situations where interpretability and explainability are critical for decision-making [38]. SVR performance can be affected by the scale of the input features. It is often necessary to scale the features appropriately to ensure that the model is not biased towards variables with larger scales [39]. The presence of multiple parameters, such as the regularization parameter ( $\alpha_i$ ) and kernel parameters, increases the complexity of hyperparameter tuning. Grid search or other optimization techniques may be required to find the optimal set of parameters [40]. SVR is designed for continuous data and may not handle categorical variables naturally. Preprocessing steps, such as one-hot encoding, may be necessary for datasets containing categorical features [41]. In cases where the kernel is too complex or the regularization parameter is too small, SVR may be prone to overfitting, capturing noise in the training data and leading to poor generalization on new data [42]. The computational and memory requirements of SVR may be challenging in resource-constrained environments, limiting its practicality in certain applications [43].

All of these advantages and disadvantages were summarized in the Table II.

### 3) Neural Network Regression:

Neural Network Regression (NNR) models (Figure 3 and equation 4), inspired by the structure of the human brain, consist of interconnected nodes (neurons) organized in layers. These models can capture complex relationships and are particularly effective in situations where the relationship between input features and SOC is intricate and non-linear [44].

$$SOC = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) \quad (4)$$

Which,  $W_1$ ,  $W_2$  are the weights,  $b_1$ ,  $b_2$  are the bias terms,  $f$  is the activation function, and  $X$  is the input features.

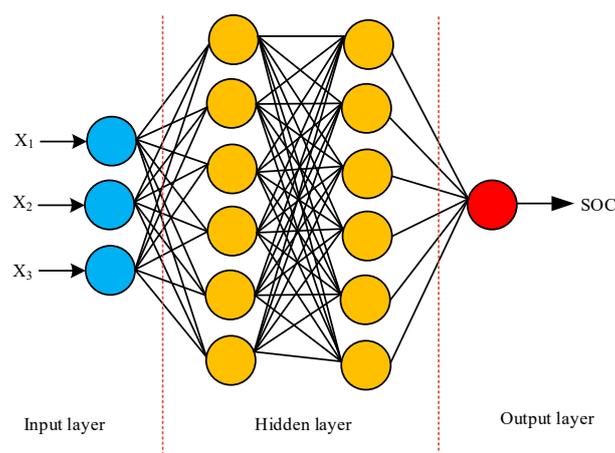


Figure 3. Neural Network Regression.

TABLE II. THE ADVANTAGES AND DISADVANTAGES OF SVR.

Ref.	Advantages	Disadvantages
[26]	Nonlinearity Handling	Computational Intensity
[27]	High-Dimensional Spaces	Model Complexity
[28]	Kernel Trick	Interpretability
[29]	Robustness to Outliers	Memory Usage
[30]	Global Optimization	Sensitivity to Noise
[31]	Flexibility in Kernel Selection	Black-Box Nature
[32]	Tuning Parameters	Data Scaling Importance
[33]	Effective in Small Sample Sizes	Large Parameter Search Space
[34]	Prediction Accuracy	Limited Handling of Categorical Data
[35]	Generalization Capability	Overfitting Risk
[36]	Resource Efficiency	Resource Requirements
[37]	Ease of Model Comparison	Limited Interpretability
[38]	Adaptability to Various Distributions	Data Preprocessing Challenges
[39]	Robust to Irrelevant Features	Limited Handling of Time-Series Data
[40]	Facilitates Hypothesis Testing	Assumption of Homoscedasticity
[41]	Versatility in Problem Types	Dependency on Kernel Choice
[42]	Ease of Hyperparameter Tuning	Limited Handling of Missing Data
[43]	Robustness to Nonlinearities	Difficulty with Non-Continuous Variables

NNR offers several advantages for estimating SOC in HEVs. Neural Networks (NNs) can model complex, nonlinear relationships between input variables (such as current, voltage, and temperature) and SOC [36]. This flexibility allows them to capture intricate patterns that may be challenging for linear models [37]. NNs can adapt to different types of data and learn representations at various levels of abstraction. This adaptability makes them suitable for diverse and dynamic SOC estimation scenarios in HEVs [38]. NNs can automatically learn relevant features from the input data. This is especially beneficial when dealing with a large number of variables or when the relationships are not explicitly known [39]. NNs can capture interactions and dependencies between variables, allowing for a more comprehensive understanding of the complex dynamics affecting SOC. This is crucial in the context of HEVs, where multiple factors influence battery state [40]. NNs can take advantage of parallel processing capabilities, which can significantly speed up training and prediction processes. This is beneficial in applications where computational efficiency is essential [41]. NNs can handle noisy data and outliers to some extent. Their ability to generalize from diverse and potentially imperfect datasets makes them robust in real-world scenarios [42]. NNs excel at representation learning, automatically discovering relevant patterns and features within the data. This is advantageous when the relationships between variables are intricate or not well-defined [43]. NNs can adapt to changes in the underlying patterns of the data over time. This adaptability is crucial in dynamic environments where the behavior of the battery or the driving conditions may evolve [44]. NNs can naturally incorporate temporal information, making them suitable for time-series data. This is essential for SOC estimation in HEVs, where battery dynamics evolve over time [45]. NNs can effectively handle large datasets, which is beneficial when training on extensive historical data to improve model accuracy [46]. NNs allow for end-to-end learning, where the model learns the mapping from input to output directly. This can simplify the modeling process and reduce the need for extensive feature engineering [47]. NNs are versatile and can be applied to various regression tasks, including SOC estimation. Their architecture can be adjusted to suit the specific requirements of the problem at hand [48].

While NNR has several advantages, it also comes with certain disadvantages that need to be considered in the context of estimating SOC in HEVs. Neural networks, especially deep architectures, can be prone to overfitting, particularly when the model complexity is high. This means the network may perform well on training data but generalize poorly to new, unseen data [29]. NNs often require a large amount of data for training to effectively capture the underlying patterns [30]. In scenarios where data is limited, the model's performance may suffer. Training NNs, especially deep architectures, can be computationally intensive. This may pose challenges in applications where real-time predictions or resource constraints are critical [32]. NNs are often considered "black-box" models, meaning it can be challenging to interpret how the model arrives at a particular prediction. This lack of transparency can be a drawback in applications where interpretability is crucial [34]. NNs have several hyperparameters (e.g., number of layers, neurons per layer, learning rate) that need to be tuned for optimal performance. Finding the right combination can be a complex and time-consuming process [35]. The performance of NNs is highly dependent on the quality, representativeness, and cleanliness of the training data. Noisy or biased data can lead to suboptimal models [40]. NNs can be sensitive to outliers in the training data. Outliers may disproportionately influence the network's weights, affecting the overall model performance [41]. NNs may require a large number of training samples to learn complex relationships effectively. In situations with limited data, the model may struggle to generalize well [42]. In deep NNs, the gradients during backpropagation can become extremely small (vanishing gradients) or large (exploding gradients). This can hinder the training process and affect convergence [43]. NN performance can be sensitive to the choice of initial weights. Poor initialization may lead to longer training times or convergence to suboptimal solutions [45]. NNs typically provide point predictions and do not naturally capture uncertainty. In situations where uncertainty estimation is crucial, additional techniques or model architectures may be required [48]. NNs can inadvertently learn biases present in the training data. Addressing ethical concerns related to bias and fairness requires careful attention to data selection and preprocessing.

### C. Evaluation Metrics

Quantifying the performance of regression models requires the use of appropriate evaluation metrics. Common metrics include:

**(1) Mean Absolute Error (MAE):** MAE represents the average absolute difference between actual ( $y_i$ ) and predicted ( $\hat{y}_i$ ) SOC values. For this, we use equation (6) [42]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

**(2) Root Mean Squared Error (RMSE):** RMSE provides a measure of the average magnitude of the errors, giving more weight to larger errors compared to MAE. We will have [43]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2} \quad (7)$$

**Coefficient of Determination ( $R^2$ ):**  $R^2$  assesses the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher  $R^2$  value indicates better predictive performance. This metric is calculated by equation (8) [47].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

TABLE III. THE ADVANTAGES AND DISADVANTAGES OF NNR

Ref.	Advantages	Disadvantages
[36]	Nonlinear Mapping	Complexity and Overfitting
	Highly Adaptive	Data Requirement
[37]	Feature Learning	Computational Intensity
	Ability to Model Interactions	Difficulty in Interpretability
	Parallel Processing	Hyperparameter Tuning Complexity
[38]	Robustness to Noisy Data	Dependency on Quality of Data
	Capacity for Representation Learning	Vulnerability to Outliers
	Adaptation to Changes	Limited Sample Efficiency
[39]	Integration of Temporal Information	Potential for Vanishing or Exploding Gradients
[40]	Ability to Handle Large Datasets	Dependency on Initialization
[41]	End-to-End Learning	Lack of Uncertainty Estimation
[42]	Versatility	Ethical and Bias Concerns
[43]	Flexibility in Model Architecture	Limited Interpretability
[44]	Automatic Feature Extraction	Data Preprocessing Challenges
[45]	Adaptability to Dynamic Environments	Difficulty with Non-Continuous Variables
[46]	Capability for Transfer Learning	Lack of Guarantees on Convergence
[47]	Integration with Temporal Dependencies	Limited Handling of Missing Data
[48]	Robustness to Irrelevant Features.	Dependency on Batch Size

#### IV. Conclusion

The estimation of State of Charge (SOC) in Hybrid Electric Vehicles (HEVs) is a critical aspect of optimizing battery performance and enhancing overall vehicle efficiency. Various regression models, each with its distinct advantages and disadvantages, offer valuable tools for this estimation process. Linear regression, with its simplicity and interpretability, provides a baseline understanding of linear relationships within the data. Support Vector Regression (SVR) extends this capability to handle nonlinear patterns, making it suitable for complex relationships in high-dimensional spaces. On the other hand, Neural Network Regression (NNR) emerges as a powerful and adaptive approach, excelling in capturing intricate nonlinearities and adapting to diverse data types. The ability to automatically learn relevant features and model complex interactions positions NNR as a versatile tool for SOC estimation in the dynamic context of HEVs. However, the choice of regression model should be made judiciously, considering the specific requirements of the application. While NNR brings flexibility and adaptability, it comes with challenges such as computational intensity, interpretability concerns, and a higher demand for data. Linear regression and SVR, while more straightforward, may struggle with capturing complex patterns. The optimal selection of a regression model hinges on a thorough understanding of the data characteristics, the desired level of interpretability, and the computational resources available. In practice, a combination of these models or hybrid approaches may be considered for a more robust and accurate SOC estimation in HEVs.

#### References

1. A. S. Mohammed, S. M. At naw, A. O. Salau, and J. N. Eneh, "Review of optimal sizing and power management strategies for fuel cell/battery/super capacitor hybrid electric vehicles," *Energy Reports*, vol. 9, pp. 2213–2228, Dec. 2023.
2. Y. Cao, M. Yao, and X. Sun, "An Overview of Modelling and Energy Management Strategies for Hybrid Electric Vehicles," *Applied sciences*, vol. 13, no. 10, pp. 5947–5947, May 2023.
3. A. Mousaei, M. Gheisarnejad, and M.H. Khooban, "Challenges and opportunities of FACTS devices interacting with electric vehicles in distribution networks: A technological review," *Journal of energy storage*, vol. 73, pp. 108860–108860, Dec. 2023.

4. Q. Zhang, S. Tian, and X. Lin, "Recent Advances and Applications of AI-Based Mathematical Modeling in Predictive Control of Hybrid Electric Vehicle Energy Management in China," *Electronics*, vol. 12, no. 2, p. 445, Jan. 2023.
5. X. Sun, J. Fu, H. Yang, M. Xie, and J. Liu, "An energy management strategy for plug-in hybrid electric vehicles based on deep learning and improved model predictive control," *Energy*, vol. 269, p. 126772, Apr. 2023.
6. Tabbi Wilberforce, Afaaq Anser, Jangam Aishawarya Swamy, and R. Opoku, "An investigation into hybrid energy storage system control and power distribution for hybrid electric vehicles," *Energy*, vol. 279, pp. 127804–127804, Sep. 2023.
7. M. Y. Nassar, M. L. Shaltout, and H. A. Hegazi, "Multi-objective optimum energy management strategies for parallel hybrid electric vehicles: A comparative study," *Energy Conversion and Management*, vol. 277, p. 116683, Feb. 2023.
8. A. Mousaei, N. Rostami, and M. B. Bannae Sharifian, "Design a robust and optimal fuzzy logic controller to stabilize the speed of an electric vehicle in the presence of uncertainties and external disturbances," Jun. 2023.
9. L. Bartolucci, E. Cennamo, S. Cordiner, V. Mulone, F. Pasqualini, and M. A. Boot, "Digital twin of a hydrogen Fuel Cell Hybrid Electric Vehicle: Effect of the control strategy on energy efficiency," *International Journal of Hydrogen Energy*, Dec. 2022.
10. Y. Liu, Y. Wu, X. Wang, L. Li, Y. Zhang, and Z. Chen, "Energy management for hybrid electric vehicles based on imitation reinforcement learning," *Energy*, vol. 263, p. 125890, Jan. 2023.
11. A. Mousaei and Y. Naderi, "Optimal Predictive Torque Distribution Control System to Enhance Stability and Energy Efficiency in Electric Vehicles," *Sustainability*, vol. 15, no. 20, pp. 15155–15155, Oct. 2023.
12. Y. Huang, H. Hu, J. Tan, C. Lu, and D. Xuan, "Deep reinforcement learning based energy management strategy for range extend fuel cell hybrid electric vehicle," *Energy Conversion and Management*, vol. 277, p. 116678, Feb. 2023.
13. Y. Wang, Y. Wu, Y. Tang, Q. Li, and H. He, "Cooperative energy management and eco-driving of plug-in hybrid electric vehicle via multi-agent reinforcement learning," *Applied Energy*, vol. 332, pp. 120563–120563, Feb. 2023.
14. C. Wu, J. Ruan, H. Cui, B. Zhang, T. Li, and K. Zhang, "The application of machine learning based energy management strategy in multi-mode plug-in hybrid electric vehicle, part I: Twin Delayed Deep Deterministic Policy Gradient algorithm design for hybrid mode," *Energy*, vol. 262, p. 125084, Jan. 2023.
15. P. Mei, Hamid Reza Karimi, H.-H. Xie, F. Chen, C. Huang, and S. Yang, "A deep reinforcement learning approach to energy management control with connected information for hybrid electric vehicles," *Engineering Applications of Artificial Intelligence*, vol. 123, pp. 106239–106239, Aug. 2023.
16. H. Wang, Y. Ye, J. Zhang, and B. Xu, "A comparative study of 13 deep reinforcement learning based energy management methods for a hybrid electric vehicle," *Energy*, vol. 266, p. 126497, Mar. 2023.
17. Pedro Maroto Estrada, Daniela, P. Bauer, M. Mammetti, and Joan Carles Bruno, "Deep learning in the development of energy Management strategies of hybrid electric Vehicles: A hybrid modeling approach," *Applied Energy*, vol. 329, pp. 120231–120231, Jan. 2023.
18. N. Yang, S. Ruan, L. Han, H. Liu, L. Guo, and X. Chen, "Reinforcement learning-based real-time intelligent energy management for hybrid electric vehicles in a model predictive control framework," *Energy*, vol. 270, pp. 126971–126971, May 2023.
19. Arash Mousaei and H. Peng, "A new control method for the steadiness of electric vehicles with 2-motor in rear and front wheels," *International Journal of Emerging Electric Power Systems*, vol. 0, no. 0, Jul. 2023.
20. Amr El-Desoky Mousa, "Extended-deep Q-network: A functional reinforcement learning-based energy management strategy for plug-in hybrid electric vehicles," *Engineering Science and Technology, an International Journal*, vol. 43, pp. 101434–101434, Jul. 2023.
21. Y. Ye, J. Zhang, S. Pilla, A. M. Rao, and B. Xu, "Application of a new type of lithium-sulfur battery and reinforcement learning in plug-in hybrid electric vehicle energy management," *Journal of Energy Storage*, vol. 59, p. 106546, Mar. 2023.
22. Z. Zhang, T. Zhang, J. Hong, H. Zhang, and J. Yang, "Energy management strategy of a novel parallel electric-hydraulic hybrid electric vehicle based on deep reinforcement learning and entropy evaluation," *Journal of Cleaner Production*, vol. 403, pp. 136800–136800, Jun. 2023.
23. N. Yang, L. Han, R. Li, Z. Wei, H. Liu, and C. Xiang, "Multiobjective Intelligent Energy Management for Hybrid Electric Vehicles Based on Multiagent Reinforcement Learning," *IEEE Transactions on Transportation Electrification*, vol. 9, no. 3, pp. 4294–4305, Sep. 2023.
24. N. Yang, L. Han, R. Li, Z. Wei, H. Liu, and C. Xiang, "Multiobjective Intelligent Energy Management for Hybrid Electric Vehicles Based on Multiagent Reinforcement Learning," *IEEE Transactions on Transportation Electrification*, vol. 9, no. 3, pp. 4294–4305, Sep. 2023.
25. C. Wang, Q. Li, A. Tang, and Z. Zhang, "A comparative study of state of charge estimation methods of ultracapacitors for electric vehicles considering temperature characteristics," *Journal of Energy Storage*, vol. 63, pp. 106908–106908, Jul. 2023.
26. J. Ruan *et al.*, "The application of machine learning-based energy management strategy in a multi-mode plug-in hybrid electric vehicle, part II: Deep deterministic policy gradient algorithm design for electric mode," *Energy*, vol. 269, pp. 126792–126792, Apr. 2023.
27. H. Zhang, J. Peng, H. Dong, H. Tan, and F. Ding, "Hierarchical reinforcement learning based energy management strategy of plug-in hybrid electric vehicle for ecological car-following process," *Applied Energy*, vol. 333, p. 120599, Mar. 2023.
28. J. Guo, J. Wang, X. Qing, B. Wang, and K. Li, "Deep Reinforcement Learning-based Hierarchical Energy Control Strategy of a Platoon of Connected Hybrid Electric Vehicles through Cloud Platform," *IEEE Transactions on Transportation Electrification*, pp. 1–1, Jan. 2023.

29. H. Lu, F. Tao, Z. Fu, and H. Sun, "Battery-degradation-involved energy management strategy based on deep reinforcement learning for fuel cell/battery/ultracapacitor hybrid electric vehicle," *Electric Power Systems Research*, vol. 220, p. 109235, Jul. 2023.
30. H. Chen, G. Guo, B. Tang, G. Hu, X. Tang, and T. Li, "Data-driven transferred energy management strategy for hybrid electric vehicles via deep reinforcement learning," *Energy Reports*, vol. 10, pp. 2680–2692, Nov. 2023.
31. L. Guo, Z. Li, Rachid Outbib, and F. Gao, "Function approximation reinforcement learning of energy management with the fuzzy REINFORCE for fuel cell hybrid electric vehicles," *Energy and AI*, vol. 13, pp. 100246–100246, Jul. 2023.
32. B. Hu, S. Zhang, and B. Liu, "A hybrid algorithm combining data-driven and simulation-based reinforcement learning approaches to energy management of hybrid electric vehicles," *IEEE Transactions on Transportation Electrification*, pp. 1–1, Jan. 2023.
33. T. Mazhar *et al.*, "Electric Vehicle Charging System in the Smart Grid Using Different Machine Learning Methods," *Sustainability*, vol. 15, no. 3, p. 2603, Feb. 2023.
34. V. S. R. Kosuru and A. Kavasseri Venkitaraman, "A Smart Battery Management System for Electric Vehicles Using Deep Learning-Based Sensor Fault Detection," *World Electric Vehicle Journal*, vol. 14, no. 4, p. 101, Apr. 2023.
35. Z. Ma, Q. Huo, W. Wang, and T. Zhang, "Voltage-temperature aware thermal runaway alarming framework for electric vehicles via deep learning with attention mechanism in time-frequency domain," *Energy*, vol. 278, pp. 127747–127747, Sep. 2023.
36. D. Qiu, Y. Wang, W. Hua, and G. Strbac, "Reinforcement learning for electric vehicle applications in power systems: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 173, p. 113052, Mar. 2023.
37. A. Alshammari and R. C. Chabaan, "Metaheuristic Optimization Based Ensemble Machine Learning Model for Designing Detection Coil with Prediction of Electric Vehicle Charging Time," *Sustainability*, vol. 15, no. 8, p. 6684, Apr. 2023.
38. D. Liu, P. Zeng, S. Cui, and C. Song, "Deep Reinforcement Learning for Charging Scheduling of Electric Vehicles Considering Distribution Network Voltage Stability," *Sensors*, vol. 23, no. 3, p. 1618, Feb. 2023.
39. H. Sun, F. Tao, Z. Fu, A. Gao, and H. Sun, "Driving-Behavior-Aware Optimal Energy Management Strategy for Multi-Source Fuel Cell Hybrid Electric Vehicles Based on Adaptive Soft Deep-Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4127–4146, Apr. 2023.
40. N. Cui, W. Cui, and Y. Shi, "Deep Reinforcement Learning Based PHEV Energy Management with Co-Recognition for Traffic Condition and Driving Style," *IEEE transactions on intelligent vehicles*, vol. 8, no. 4, pp. 3026–3039, Apr. 2023.
41. B. P. Adedeji, "A Multivariable Output Neural Network Approach for Simulation of Plug-in Hybrid Electric Vehicle Fuel Consumption," *Green Energy and Intelligent Transportation*, p. 100070, Feb. 2023.
42. F. Millo, L. Rolando, L. Tresca, and L. Pulvirenti, "Development of a neural network-based energy management system for a plug-in hybrid electric vehicle," *Transportation Engineering*, vol. 11, pp. 100156–100156, Mar. 2023.
43. T. Dong, Z. Zhang, H. Liao, J. Pan, and Y. Xiao, "Prediction of cold start emissions for hybrid electric vehicles based on genetic algorithms and neural networks," *Journal of Cleaner Production*, vol. 420, pp. 138403–138403, Sep. 2023.
44. F. Guo, Z. Chen, F. Xiao, A. Li, and J. Shi, "Real-Time Energy Performance Benchmarking of Electric Vehicle Air Conditioning Systems Using Adaptive Neural Network and Gaussian Process Regression," *Applied Thermal Engineering*, p. 119931, Dec. 2022.
45. M. Mokhtari Mehmandoosti and F. Kowsary, "Artificial neural network-based multi-objective optimization of cooling of lithium-ion batteries used in electric vehicles utilizing pulsating coolant flow," *Applied Thermal Engineering*, vol. 219, p. 119385, Jan. 2023.
46. Y. Wang, W. Li, Z. Liu, and L. Li, "An Energy Management Strategy for Hybrid Energy Storage System Based on Reinforcement Learning," *World Electric Vehicle Journal*, vol. 14, no. 3, p. 57, Mar. 2023.
47. Bukola Peter Adedeji and G. Kabir, "A feedforward deep neural network for predicting the state-of-charge of lithium-ion battery in electric vehicles," *Decision Analytics Journal*, vol. 8, pp. 100255–100255, Sep. 2023.
48. B. P. Adedeji, "Electric vehicles survey and a multifunctional artificial neural network for predicting energy consumption in all-electric vehicles," *Results in Engineering*, vol. 19, p. 101283, Sep. 2023.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.