Synthetic-to-Smart: Enhancing EV Battery Insights Using TimeGAN and Temporal Fusion Transformer

Shashank G[1], Shivaprasad[2], Vinisha A[3], Hemanth Kumar D[4], Mrs. Divyaraj G N[5]
[1,2,3,4] *Artificial Intelligence and Machine Learning Machine Learning NITTE (Deemed to be University) Bangalore, Karnataka, India*
[5] *Assistant Professor, Dept. of AIML NITTE (Deemed to be University) Bangalore, Karnataka, India*

Abstract*—Electric vehicles (EVs) are changing the contempo-rary transportation, yet the mass utilization of the vehicles is largely conditional upon the trustworthy functioning of battery systems and their correct monitoring. The ability to predict the most important battery health indicators: State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL) is a problematic issue because the real-world data is usually not accessible, and such data is usually noisy and imbalanced. Con-ventional methods of battery degradation analysis find it difficult to reflect the nonlinear, time-varying, and complicated nature of battery behavior. This work suggests a hybrid framework, which integrates Time-series Generative Adversarial Networks (TimeGAN) to generate synthetic data with a Temporal Fu-sion Transformer (TFT) to predictive model to overcome these difficulties. TimeGAN is a neural network that is utilized to create realistic multivariate time-series data that both maintains temporal patterns and feature-relationships, and it can overcome the problem of data scarcity and distribution. The t-SNE visual-ization and statistical similarity measure is used to evaluate the quality of the generated data and make sure that it is consistent with actual battery signals. The TFT model is then trained on the augmented dataset and it utilizes attention mechanisms and gated architectures to learn long-term dependencies and nonlinear patterns of degradation. Consequently, the model can make correct predictions of SOC, SOH and RUL. On the whole, the offered framework increases the accuracy of prediction, as well as the reliability of battery monitoring systems.

***Index Terms*—Battery management system, TimeGAN, Tempo-ral fusion transformer, State of charge, State of health, Remaining useful life, Synthetic data generation**

I. INTRODUCTION

A. Background and Motivation

Electric Vehicles (EVs) have been swiftly transforming transportation in the world by providing better energy effi-ciency, decreasing the emission, and a high relationship with sustainable energy objectives. The lithium-ion battery is the core component of any EV, and its performance, reliability, and life cycle have a direct effect on the overall efficiency, safety and user-experience of the vehicle. With the ever-growing adoption of EVs, sophisticated Battery Management Systems (BMS) are now of higher significance to monitor key battery parameters, including State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL)

B. Limitations of Existing Methods

Estimation of State of Charge (SOC) and State of Health (SOH) has commonly been done by traditional battery esti-mation techniques, such as Coulomb Counting, Open Circuit Voltage (OCV)-based and Equivalent Circuit Models (ECMs). Nonetheless, they tend to fail to be consistent in various driving scenarios, changing environmental conditions, and degradation states among others

Time-Series Generative Adversarial Networks (TimeGAN) are suggested to fill the gap of data scarcity by creating syn-thetic time-series data. TimeGAN can maintain time dynamics and generate realistic sequences, which can be applied to augment battery datasets

C. Research Contributions

To deal with the abovementioned challenges, the proposed work suggests a hybrid synthetic-to-smart architecture of elec-tric vehicle (EV) battery prediction that integrates TimeGAN-based data augmentation with a Temporal Fusion Transformer (TFT) to predict multivariate time-series. The main concept is to augment small real-world data sets with high-quality synthetic data and to use advanced sequence modeling to achieve better prediction results.

TimeGAN is used to create synthetic sequence of the bat-tery, maintaining the temporal relationship and the relationship between features. To guarantee the credibility of the obtained data, it is checked with the help of t-SNE visualization, correlation analysis, and statistical similarity measures, which ensures that the data is consistent with the actual behavior of the battery.

The key findings of this study can be summarized as follows:

Hybrid TimeGAN–TFT Framework: Cohesive architecture with synthetic data generation and attention-based predictive model to achieve precise prediction of SOC, SOH and RUL. Quality Synthetic Data Generation and Testing: The resulting data has a temporal organization and relationship of features and is strictly checked to be realistic and useful. Improved Accuracy of Prediction and Strength: The framework enhances model generalization, curbs overfitting, and boosts the per-formance of multi-horizon prediction because it uses limited datasets with established synthetic samples to augment them. Scalable and Interpretable Architecture: The interpretability that is facilitated by the TFT allows attention-based inter-pretability to facilitate clear decision-making in real-world battery management systems at vehicle and fleet scales.

II. RELATED WORK

Recent developments in electric vehicle (EV) battery ana-lytics have been concerned with enhancing the accuracy of battery state estimation using deep learning, hybrid models, and synthetic data augmentation. In this section, the previous literature is reviewed in three broad groups, namely, tradi-tional data-driven methods, hybrid physical-data models, and synthetic data augmentation techniques.

A. Deep Learning-Based Battery State Estimation

State of Charge (SOC) and State of Health (SOH) prediction have been receiving substantial momentum in deep learning architectures, due to its ability to represent nonlinear relation-ships in battery behaviour. Saneep et al. Bhattacharyya et al. [8] utilized Convolutional Neural Net-works (CNNs) for estimating State of Charge (SOC) under the driving conditions of New York City. Their results indicated that while CNNs are adept at capturing local temporal features, they face challenges in modeling long-term dependencies over prolonged degradation cycles.

Jayasree et al. [6] combined the deep neural network model with self-attention modules for simultaneous State of Charge (SOC) and State of Health (SOH) prediction. This self-attention module enabled the model to focus on key degradation variables, thus, improving robustness and inter-pretability of the model. Similarly, CV et al. [9] evaluated energy prediction in electric vehicles (EV) using transformer-based tabular learning models, such as TabTransformer and TabNet. They found that the model performed well (R^2 = 0.9812), highlights the potential of attention-based models to capture complex behaviours of electric vehicles.

Despite their predictive capabilities, deep learning models necessitate large and diverse datasets to avoid overfitting and ensure generalization, as highlighted in the comprehensive review by El Fallah et al. [7]. This reliance on extensive datasets continues to pose a significant practical limitation in the analytics of electric vehicle batteries.

B. Hybrid Physical–Data Modeling Approaches

Hybrid models combine physical insights of battery systems with learning models to improve prediction accuracy. Tang et al. [3] developed hybrid state of charge (SOC) estimation method that integrates an Equivalent Circuit Model (ECM) with a Gated Recurrent Unit (GRU) network. The ECM models voltage variations using a second-order RC circuit, and the GRU captures nonlinear electrochemical effects. The proposed weighted error fusion strategy effectively improved the accuracy of the proposed hybrid approach, highlighting the benefits of blending expert knowledge and neural networks.

While hybrid models boost robustness, they still face issues with sensitivity, computational efficiency, and generalization across different operating conditions. Further, these models may need to be retrained for different battery types or envi-ronmental conditions.

The need for more real-world battery data has led to an interest in generating synthetic data. Mohanty

and colleagues
[1] proposed a TimeGAN-based data augmentation technique for State of Charge (SOC) prediction, and found that synthetic data boosted the performance of transformer models under unseen scenarios. However, their study indicated the need for more sophisticated multivariate datasets with vehicle dynam-ics.
Seol and Kim [4] also investigated the impact on TimeGAN of different preprocessing methods. Using both discriminative and predictive scores and t-SNE visualisation, they demon-strated that Max-Abs transformation and normalization en-hanced the realism of synthetic data up to 53.7
Li and Li [5] proposed a combined TCNN-Transformer-TimeGAN framework to overcome the imbalanced data of abnormal conditions. Their approach produced low ARMSE values, suggesting that the augmentation of synthetic data and attention-based models improve the degradation-based model performance.
In summary, these studies indicate that synthetic data gener-ation capability enhances model performance and generaliza-tion. But there is a lack of research on a generic architecture that combines the benefits of proven TimeGAN-based feature engineering with sophisticated prediction models such as the Temporal Fusion Transformer (TFT) to improve the joint pre-diction of SOC, SOH and RUL. This study's hybrid framework is motivated by this gap in the literature.

III. PROPOSED METHODOLOGY

Our approach aims to improve the quality of electric vehicle battery performance monitoring and analysis through the syn-thesis of datasets and forward-looking forecasting algorithms. In particular, TimeGAN is employed to create realistic simu-lations of time-series data representing the temporal dynamics of the battery. This expanded set of features is then leveraged to train a Temporal Fusion Transformer (TFT) model to generate accurate multiple-step ahead predictions of battery health metrics, such as State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL).

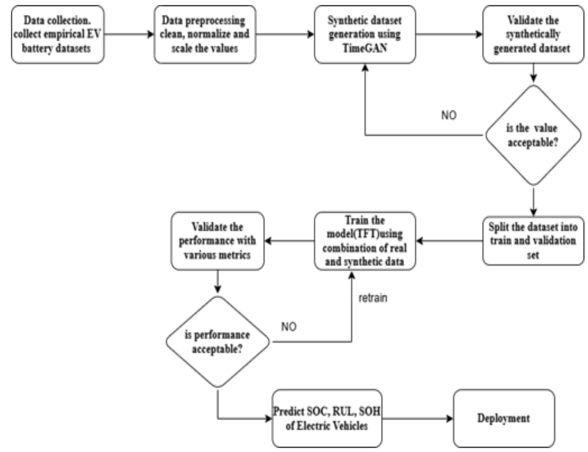

Fig. 1. Overall workflow of the proposed TimeGAN–TFT framework.

A. Dataset Description

The effectiveness of the proposed approach is demonstrated on a multivariant electric vehicle battery telemetry data that consists of voltage, current, temperature, capacity, energy, cycle index and time information. These features represent the temporal charge/discharge pattern of the battery under various working conditions. The dataset is divided into a fixed-size temporal sequence of features that are used to predict metrics representing battery health, State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL). One issue with real battery data is the distribution of the degradation process. Often, early-cycle measurements are frequent, but the measurements of later stages of degradation are scarce. This can impact the effectiveness of model training and raises the need of data augmentation tools such as synthetic data generation.

B. Data Preprocessing

The effectiveness of the proposed method is tested on a multivariate battery telemetry data matrix derived from electric vehicle batteries, which contains variables such as battery voltage, current, temperature, capacity, energy, cycle count and time-related variables. These features capture the charge and discharge dynamics of the battery including operation conditions. The data is segmented into a fixed-length time window and key battery health metrics - State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL), are defined as target predictions of interest. One of the issues when dealing with real battery datasets is the lack of information about the different stages of degradation. We often have significant data for the early part of battery cycles, but lack data describing the later phases. This can be a problem in terms of model generalization, and underlines the importance of data augmentation, including the generation of synthetic data.

C. TimeGAN-Based Synthetic Data Generation

In order to overcome the issues arising from data scarcity and imbalance, Time-series Generative Adversarial Networks (TimeGAN) are employed to produce realistic, multivariate sequences of battery data. The model is trained with a combination of reconstruction loss, supervised loss, and an adversarial loss:

$$L_{total} = L_{recon} + \alpha L_{sup} + \beta L_{adv} \quad (1)$$

This joint loss function follows the approach of the original TimeGAN paper. [11].

D. Temporal Fusion Transformer (TFT) Architecture

Our model is the Temporal Fusion Transformer (TFT) from PyTorch Forecasting, which predicts the State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL) for multiple time steps. The model comprises approx-imately 19,000 parameters, allowing it to be efficient yet powerful to learn complex structures. The model consists of:

- Interpretable Multi-Head Attention: This layer allows for the model to focus on long-range dependencies and is able to weigh different temporal steps.
- Position-wise Feed Forward Network: A gated feed-forward network is used to improve the features before the prediction.
- Final Linear Layer: This fully-connected layer outputs the prediction for SOC, SOH and RUL.

The attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

This structure is borrowed from the Transformer model [12], on which the TFT model is based. Combining recurrent and attention mechanisms allows the model to effectively capture both short and long term changes in battery behavior, while allowing some degree of interpretability.

E. Model training and loss function

The Temporal Fusion Transformer (TFT) model is based on PyTorch Lightning and uses GPUs for training. Training occurs for up to 30 epochs with early stopping to avoid overfitting. A gradient clip of 0.1 is used to keep the gradients stable. Adam optimizer is used to optimise the model.

Being regression problems, the State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL) prediction are done using the RMSE loss function:

$$L_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (3)$$

where y_i represents the true value and $\hat{y}_i$ denotes the predicted value. RMSE is used to penalise larger losses - this is crucial in the prediction of long-term battery degradation.

F. Health Indicator Formulation

The State of Charge (SOC) is defined as:

$$SOC = \frac{Q_{\text{remaining}}}{Q_{\text{rated}}} \times 100 \quad (4)$$

The State of Health (SOH) is defined as:

$$SOH = \frac{C_{\text{current}}}{C_{\text{rated}}} \times 100 \quad (5)$$

The Remaining Useful Life (RUL) is calculated as:

$$RUL = \text{MaxCycle} - \text{CurrentCycle} \quad (6)$$

IV. RESULTS AND DISCUSSION

A. TimeGAN Performance and Synthetic Data Validation

The Time-series Generative Adversarial Network (TimeGAN) was trained to generate feasible multivariate time series of electric vehicle (EV) batteries. To observe the stability of the training process, we monitored the reconstruction, supervisory, generator and discriminator losses. The convergence values for the losses considered are tabulated in Table I. The minimal reconstruction and

TABLE I FINAL TIMEGAN LOSS VALUES AFTER CONVERGENCE

Loss Component	Value
Embedder Loss	0.0081
Supervisor Loss	0.0111

Reconstruction Loss	0.0004
Generator Loss	2.6850
Discriminator Loss	0.6528

supervised losses suggest that the model preserved the dynamic and step-by-step relationships of the sequences. The low losses of the generator and discriminator indicate that the adversarial training was effective, with no signs of mode collapse. To visually inspect the distributions of the generated sequences, we used t-SNE to plot the real and synthetic data in a lower-dimensional space. In Fig. 2, we show that the synthetic data is highly overlapped with the real data in the latent space. This confirms that TimeGAN is successful at learning complex degradation process and it produces synthetic battery telemetry that can be used for forecasting.

The findings show strong forecasting capability for all three battery health indicators. The small RMSE values suggest

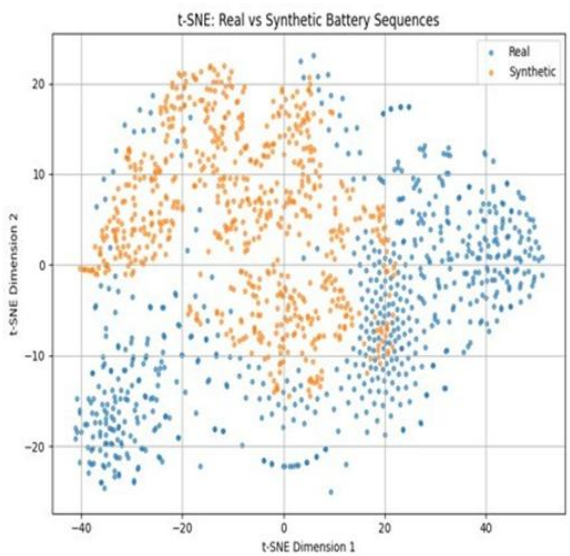

Fig. 2. t-SNE plot comparing real and synthetic battery sequences generated by TimeGAN.

TABLE II PREDICTION PERFORMANCE (RMSE, UNITS)

Target Variable	RMSE
State of Charge (SOC, %)	0.47560
State of Health (SOH, %)	0.37500
Remaining Useful Life (RUL, cycles)	0.04012

the model can capture both short-term variations in State of Charge (SOC) and long-term degradation patterns of State of Health (SOH) and Remaining Useful Life (RUL). The plots in Figures 3–5 show a visual comparison between the actual and predicted SOC, SOH and RUL curves. The prediction curves well align with the actual data, validating the effectiveness of the TFT model, which excellently captures the nonlinear battery dynamics and long-term degradation. There are slight differences during fast transition periods, which is expected due to the complex nature of the process which takes place during battery degradation. Overall, the RMSE values and the congruence of the graphs confirm the success of the TimeGAN-augmented TFT approach.

V. CONCLUSION AND FUTURE WORK

This paper proposes a synergy of deep learning models using TimeGAN-based data synthesis with a Temporal Fusion Transformer (TFT) to foretell the State of Charge (SOC), State of Health (SOH) and Remaining Useful Life (RUL) of electric vehicle batteries. The use of synthetic sequence augmentation overcomes the issue of data imbalance across different stages of battery degradation, improving the model's performance and generalization. The TFT network effectively leverages temporal information and automatically selects key features for multi-horizon forecasts with stable and consistent outcomes. Results suggest that the generated synthetic data effectively

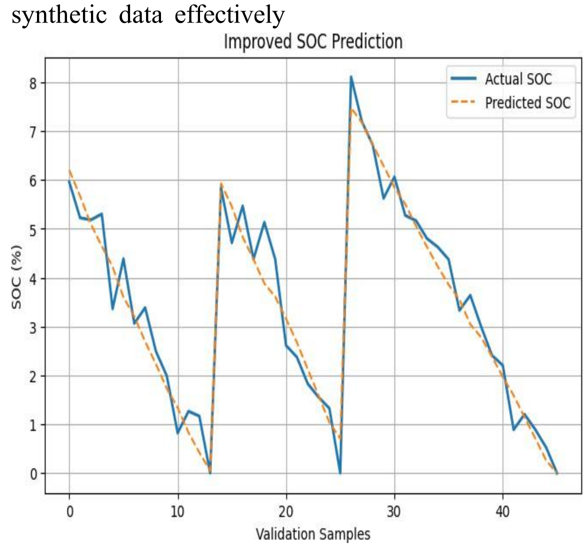

Fig. 3. Actual vs Predicted State of Charge (SOC)

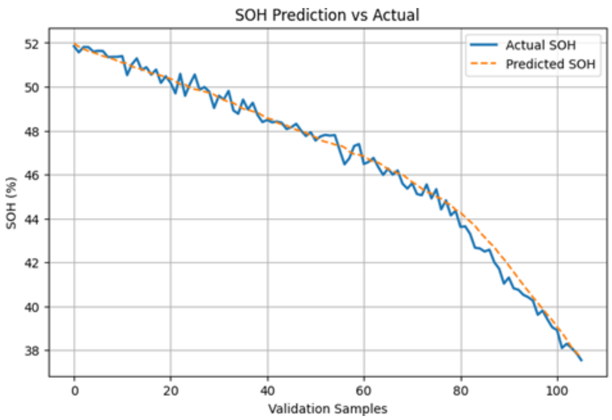

Fig. 4. Actual vs Predicted State of Health (SOH)

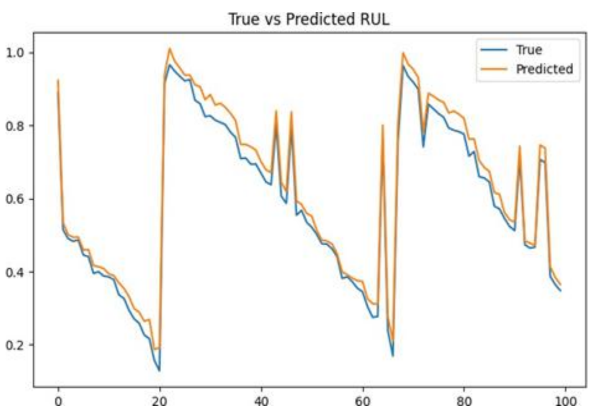

Fig. 5. Actual vs Predicted Remaining Useful Life (RUL)

retains key temporal features of real battery sequences and en-hances the forecasting accuracy. Also, uncertainty estimation methods may be used to provide uncertainty bounds for SOC, SOH and RUL predictions, which can be useful for safety-critical systems. Finally, real-time implementation in embed-ded battery management systems (BMS) or edge computing would be a major step towards industrial deployment.

ACKNOWLEDGMENT

This section has been removed for the double-blind review

REFERENCES

[1] P. K. Mohanty, P. Jena, and N. P. Padhy, "TimeGAN-based Diversified Synthetic Data Generation Following BERT-based Model for EV Battery SOC Prediction: A State-of-the-Art Approach," *IEEE Transactions on Industry Applications*, 2025.

[2] K. Saneep, M. Kannan, K. Sundareswaran, P. S. R. Nayak, and T. Mithun, "Performance Comparison of Recurrent Neural Network Variants for State of Charge Prediction in Lithium-Ion Batteries," in *2025 International Conference on Power Electronics Converters for Transportation and Energy Applications (PECTEA)*, June 2025, pp. 1–5.

[3] A. Tang, Y. Huang, S. Liu, Q. Yu, W. Shen, and R. Xiong, "A novel lithium-ion battery state of charge estimation method based on the fusion of neural network and equivalent circuit models," *Applied Energy*, vol. 348, p. 121578, 2023.

[4] S. Seol and B. Kim, "Study on Improving Reliability of Synthetic Battery Data with TimeGAN Pre-processing," in *2023 IEEE 6th Inter-national Conference on Knowledge Innovation and Invention (ICKII)*, Aug. 2023, pp. 747–750.

[5] P. Li and M. Li, "State-of-health Monitoring of Li-ion Battery Driven by Convolutional Neural Network and Attention Mechanism," in *2023 35th Chinese Control and Decision Conference (CCDC)*, May 2023, pp. 1319–1324.

[6] M. Jayasree, A. P. Darpan, R. Gopalakrishnan, V. Indumathi, and R. Radhamani, "A Hybrid Deep Neural Network and Self-Attention Mechanism for Accurate Estimation of Battery SoC and SoH," in *2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications (AIMLA)*, Apr. 2025, pp. 1–6.

[7] S. El Fallah et al., "Advanced Battery Management for Electric Vehicles: A Deep Dive into Estimation Techniques Based on Deep Learning for the State of Health and State of Charge of Lithium-Ion Batteries," in *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, Apr. 2024, pp. 1–5.

[8] H. S. Bhattacharyya, A. Yadav, A. B. Choudhury, and C. K. Chanda, "Convolution neural network-based SOC estimation of Li-ion battery in EV applications," in *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques*

(ICEECCOT), Dec. 2021, pp. 587–592.

[9] M. S. CV, M. K. Amal, R. Satheesh, and H. H. Alhelou, “Enhanced Electric Vehicle Energy Consumption Prediction With TabTransformer, TabNet, and Bidirectional Encoder Representations From Transformers Embeddings,” *IEEE Transactions on Industrial Informatics*, 2025.

[10] S. Seol, J. Lee, J. Yoon, and B. Kim, “Improving SOH estimation for lithium-ion batteries using TimeGAN,” *Machine Learning: Science and Technology*, vol. 4, no. 4, p. 045007, 2023.

[11] J. Yoon, D. Jarrett, and M. van der Schaar, “Time-series Generative Adversarial Networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[12] A. Vaswani et al., “Attention Is All You Need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.