

Hyperscale-Cascaded Transformer-Net-Based Framework for Remaining Useful Life Prediction of Electric Vehicle Batteries

¹Siva Ganesh Chinthakula, ²G. Satya Narayana, ¹D. Indu, ¹CH. Raghava Reddy, ¹G. Lahari

¹Godavari Institute of Engineering and Technology (A), Rajahmundry, India

²Godavari Global University, Rajahmundry, India

Abstract. The accurate prediction of remaining useful life (RUL) of electric vehicle (EV) batteries is a critical aspect of intelligent battery management systems. Effective RUL prediction not only ensures vehicle safety and reliability but also plays a pivotal role in optimizing charging cycles, reducing maintenance costs, and extending the overall battery lifespan. This work presents a comprehensive Deep Learning (DL) framework for predicting RUL of EV batteries, using a novel Hyperscale-Cascaded Transformer Net architecture designed to capture long-term dependencies and degradation patterns in battery behavior. The proposed system initiates with data acquisition, wherein parameters such as cycle index, voltage, current, and time-based features are collected. Raw data undergoes preprocessing, which includes data cleaning to eliminate outliers and handle missing values, followed by Exploratory Data Analysis (EDA) to extract meaningful patterns through descriptive statistics, distribution analysis, and correlation heatmaps. Subsequently, the data is passed through a feature engineering pipeline, where feature scaling using Min-Max normalization is applied to enhance learning efficiency of model. Processed dataset is then split into training and testing sets, maintaining data integrity for unbiased evaluation. The core of the model lies in Hyperscale-Cascaded Transformer Net, a DL model that utilizes cascaded transformer layers to model complex temporal relationships and nonlinear degradation behaviors inherent in battery performance over time. The model validation and performance evaluation are conducted using Python software, and performance metrics are measured in terms of error metrics such as Mean Absolute Error (MAE) of 0.0211, Mean Square Error (MSE) of 0.0006, Root Mean Squared Error (RMSE) of 0.0245, and coefficient of determination (R^2 -score) of 0.9993. Experimental results demonstrate that proposed Transformer-based model outperforms traditional Machine Learning (ML) techniques in terms of accuracy and robustness in revolutionizing EV battery management systems.

Keywords: RUL prediction, EV batteries, hyperscale-cascaded Transformer Net architecture, exploratory data analysis, Python software.

DOI: <https://doi.org/10.52254/1857-0070.2026.1-69.10>

UDC: 621.355

Infrastructură cascadă hiper-scalabilă bazată pe Transformernet pentru estimarea duratei de viață reziduale a bateriilor vehiculelor electrice

¹Siva Ganesh Chinthakula, ²G. Satya Narayana, ¹D Indu, ¹CH Raghava Reddy, ¹G Lahari

¹Profesor asistent, Departamentul de Inginerie Electrică și Electronică, Institutul de Inginerie și Tehnologie Godavari (A), Rajahmundry, India

²Profesor, Departamentul de Inginerie Electrică și Electronică, Universitatea Globală Godavari, Rajahmundry, India

Rezumat. Predicția precisă a duratei de viață utilă rămase (RUL) a bateriilor vehiculelor electrice (EV) este un aspect critic al sistemelor inteligente de gestionare a bateriilor. Predicția eficientă a duratei de viață utilă rămase (RUL) nu numai că asigură siguranța și fiabilitatea vehiculului, dar joacă și un rol esențial în optimizarea ciclurilor de încărcare, reducerea costurilor de întreținere și extinderea duratei de viață totale a bateriei. Această lucrare prezintă un cadru cuprinzător de Deep Learning (DL) pentru prezicerea duratei de viață utilă rămase (RUL) a bateriilor EV, utilizând o nouă arhitectură TransformerNet Hyperscale-Cascaded, concepută pentru a capta dependențele pe termen lung și modelele de degradare în comportamentul bateriei. Sistemul propus inițiază cu achiziția de date, în care sunt colectați parametri precum indicele ciclului, tensiunea, curentul și caracteristicile bazate pe timp. Datele brute sunt supuse preprocesării, care include curățarea datelor pentru a elimina valorile aberante și a gestiona valorile lipsă, urmată de analiza exploratorie a datelor (EDA) pentru a extrage modele semnificative prin statistici descriptive, analiză a distribuției și hărți termice de corelație. Ulterior, datele sunt transmise printr-o conductă de inginerie a caracteristicilor, unde se aplică scalarea caracteristicilor folosind normalizarea Min-Max pentru a îmbunătăți eficiența învățării modelului. Setul de date procesat este apoi împărțit

în seturi de antrenament și testare, menținând integritatea datelor pentru o evaluare imparțială. Nucleul modelului constă în TransformerNet în cascadă hiperscală, un model DL care utilizează straturi de transformare în cascadă pentru a modela relații temporale complexe și comportamente de degradare neliniară inerente performanței bateriei în timp. Validarea modelului și evaluarea performanței sunt efectuate folosind software-ul Python, iar indicatorii de performanță sunt măsurati în termeni de indicatori de eroare, cum ar fi eroarea absolută medie (MAE) de 0.062, eroarea pătratică medie (MSE) de 0.0006, eroarea pătratică medie rădăcină (RMSE) de 0.0093 și coeficientul de determinare (scor R^2) de 0.9989. Rezultatele experimentale demonstrează că modelul propus bazat pe Transformer depășește tehnicile tradiționale de învățare automată (ML) în ceea ce privește precizia și robustețea, revoluționând sistemele de gestionare a bateriilor pentru vehicule electrice.

Cuvinte cheie: predicție RUL, baterii EV, arhitectură TransformerNet în cascadă la hiperscală, EDA, software Python.

Гипермасштабная каскадная инфраструктура на основе Transformernet для прогнозирования остаточного срока службы аккумуляторов электромобилей

¹Шива Ганеш Чинтхакула, ²Г. Сатья Нараяна, ¹Д Инду, ¹Ч. Рагхава Редди, ¹Г. Лахари

¹Инженерно-технологический институт Годавари (А), Раджамандри, Индия

²Глобальный университет Годавари, Раджамандри, Индия

Аннотация. Точное прогнозирование остаточного срока службы (RUL) аккумуляторов электромобилей (ЭМ) является критически важным аспектом интеллектуальных систем управления аккумуляторами. Эффективное прогнозирование остаточного срока службы (RUL) не только обеспечивает безопасность и надежность транспортного средства, но и играет ключевую роль в оптимизации циклов зарядки, снижении затрат на техническое обслуживание и продлении общего срока службы аккумуляторов. В данной работе представлена комплексная платформа глубокого обучения (DL) для прогнозирования остаточного срока службы аккумуляторов ЭМ с использованием новой гипермасштабируемой каскадной архитектуры TransformerNet, разработанной для выявления долгосрочных зависимостей и закономерностей деградации в поведении аккумуляторов. Предлагаемая система начинается со сбора данных, в ходе которого собираются такие параметры, как индекс цикла, напряжение, ток и временные характеристики. Исходные данные проходят предварительную обработку, которая включает очистку данных для устранения выбросов и обработки пропущенных значений, после чего проводится разведочный анализ данных (EDA) для извлечения значимых закономерностей с помощью описательной статистики, анализа распределения и построения тепловых карт корреляций. Затем данные проходят через конвейер проектирования признаков, где применяется масштабирование признаков с использованием нормализации Min-Max для повышения эффективности обучения модели. Обработанный набор данных затем разделяется на обучающий и тестовый наборы, сохраняя целостность данных для беспристрастной оценки. Ядром модели является Hyperscale-Cascaded TransformerNet, модель DL, которая использует каскадные слои трансформатора для моделирования сложных временных соотношений и нелинейного поведения деградации, присущего производительности батареи с течением времени. Валидация модели и оценка производительности проводятся с помощью программного обеспечения Python, а метрики производительности измеряются в терминах метрик ошибок, таких как средняя абсолютная ошибка (MAE) 0.062, среднеквадратическая ошибка (MSE) 0.0006, среднеквадратическая ошибка (RMSE) 0.0245 и коэффициент детерминации (R^2 -score) 0.9993. Экспериментальные результаты показывают, что предлагаемая модель на основе трансформатора превосходит традиционные методы машинного обучения (ML) с точки зрения точности и надежности в революционных системах управления батареями электромобилей.

Ключевые слова: прогнозирование остаточного ресурса аккумуляторов, аккумуляторы электромобилей, гипермасштабная каскадная архитектура TransformerNet, EDA, программное обеспечение Python.

I. INTRODUCTION

Rapid developments in battery technology, including high stability, low self-discharge rate, and lightweight design, coupled with growing worries about climate change, have led to the widespread adoption of batteries in a variety of applications, most notably EVs. Despite the global growth in the hybrid and EV market, a critical factor for accelerating their adoption is improving operational efficiency [1]. Due to the

lack of precise techniques for estimating its end of life, battery packs the costliest component are frequently utilized inefficiently or replaced too soon. According to industry standards, a battery is said to be near the end of its life when its capacity drops below its initial value [2]. However, as direct on-board measurements are not practical, this capability accurately determined by laboratory testing. Consequently, Battery Management Systems (BMS) in EVs

often rely on simplistic estimations, such as counting ampere-hour throughput or discharge cycles, which fail to account for variations in driving behavior, usage intensity, and environmental conditions like climate and terrain [3].

EV users are increasingly concerned about battery's health status and its RUL, as severely degraded batteries pose safety risks, including potential fire or explosion hazards. Consequently, it is essential for BMS to monitor RUL and manage battery lifespan effectively. Battery life management plays a pivotal role in BMS operation, directly influencing evaluation and utilization of the vehicle's Battery Energy Storage System (BESS). To achieve this, it is necessary to understand performance variation among individual cells within battery pack and determine their current RUL [4, 5]. Nonetheless, battery degradation arises from complex internal electrochemical reactions, making the underlying mechanisms difficult to explain. Among these, electrochemical processes are a primary contributor to performance decline. Additionally, external parameters such as current and temperature further exacerbate degradation [6].

A more precise and adaptable method for predicting the RUL of batteries extend their service life significantly by ensuring replacement only when truly necessary [7]. It would also enhance understanding of how different operating conditions influence battery aging. Semi-empirical or empirical models for capacity loss, usually based on metrics like State of Charge (SOC) and C-rate, are frequently used in current BMS implementations. Nevertheless, these models lack stochastic components and are typically deterministic [8]. By accounting for the inherent unpredictability in aging data, a stochastic model, proposed in this study offers a more reliable estimate. Moreover, ML techniques are increasingly being explored for battery prognostics, showing promising accuracy in predicting capacity loss from operational data [9, 10].

A. Related Works

Shaheer Ansari *et al* (2021) [11] have proposed a Multi-Channel Profile (MCI) based Artificial Neural Network framework for predicting the RUL of EV lithium-ion batteries. The MCI-based ANN captures different battery degradation patterns. The model, however, has a high generalization ability when a large amount of data is used for training but its accuracy becomes

lower with a small number of training datasets, thus, its scalability is limited.

Luis Magadán *et al* (2023) [12] have implemented Bi-directional Long Short Term Memory (BiLSTM) based prediction of RUL of bearings in electric motors. It combines BiLSTM network for sequence learning and RUL prediction. However, the dependency on dataset-specific feature distributions pose a problem for the method's flexibility when it is used on motors with very non-stationary signals or for the detection of new types of bearing faults, thus, more extensions are needed for a bigger scale.

Kai Li *et al* (2020) [13] have introduced fuzzy logic (FL) in making battery life predictions more accurate for EVs with the use of charging cloud data. FL makes the prediction process more robust and accurate, thus the estimation errors are less than 4% for the whole sampled EV datasets, support the stabilization of battery life estimation even if the cloud data are uncertain and noisy. However, the requirement for predefined fuzzy rules and membership functions restrict the battery-to-battery adaptability when different chemistries or highly variable operating environments are considered.

Dexin Gao *et al* (2023) [14] have presented one Dimensional Convolutional Neural Network (CNN)-Bi-directional Long Short Term Memory (1D-CNN-BiLSTM) for forecasting RUL of lithium-ion batteries in EVs. The CNN module digs deep features from state-of-health (SOH) data in a very effective manner, whereas the BiLSTM unit gets the temporal dependencies in both the past and the future thus it can learn the sequence robustly. This hybrid model not only features the best characteristics of both worlds but requires high computational resources and extensive training data, which limit its scalability across diverse battery chemistries and operating conditions.

Contributions of the proposed EV battery RUL prediction system are given below:

- The data cleaning improves data credibility by eliminating outliers, interpolating missing values, and dampening noise to provide high-quality input for model training.
- EDA exposes latent patterns of degradation and feature correlations with statistical profiling, distribution mapping, and correlation heatmaps.
- Min-max scaling based feature engineering scales input features to a bounded range [0, 1] to enhance learning stability and ensure consistent model convergence.

The hyperscale-cascaded TransformerNet models sophisticated temporal relationships and non-linear degradation patterns with a deep, multi-block transformer framework that includes hierarchical attention.

This paper has following structure: Proposed framework's design and component parts are described in Section 2. Development and validation of RUL estimation model and capacity

loss model are covered in Section 3. Results of entire proposed framework are shown in Section 4, and the paper is concluded with a summary of results and key takeaways in Section 5.

II. PROPOSED SYSTEM DESCRIPTION

The proposed block diagram of Hyperscale cascaded TransformerNet based framework for RUL prediction in EV batteries is shown in Fig. 1.

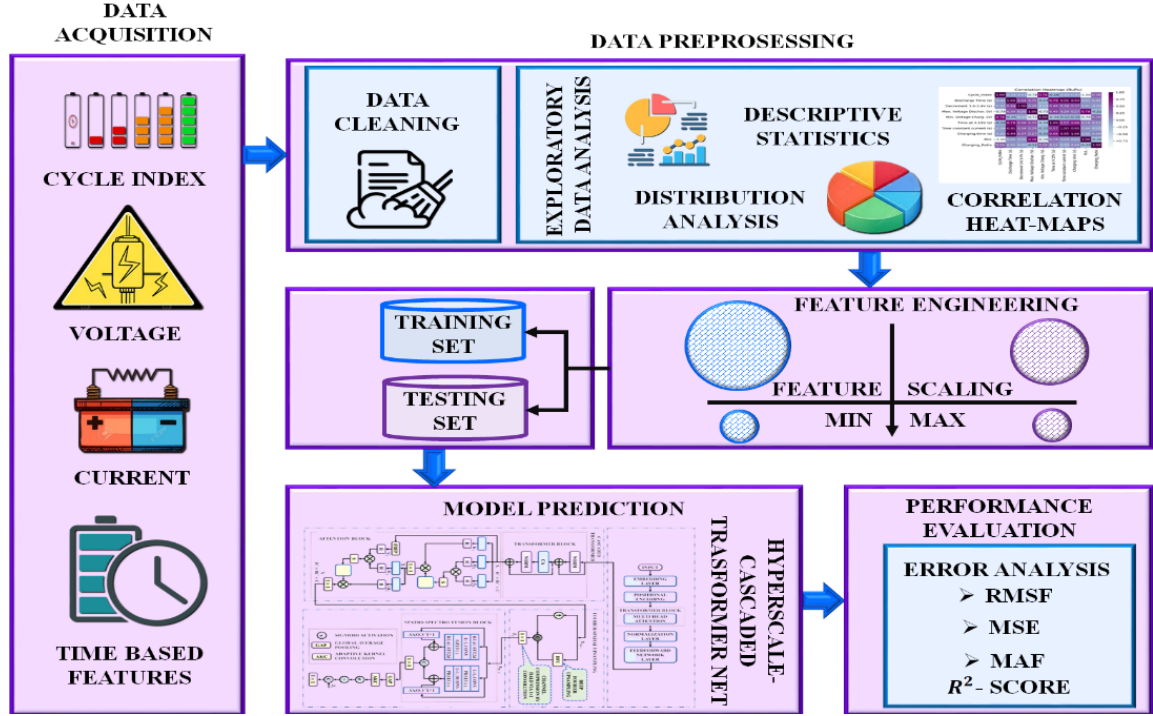


Fig. 1. Proposed RUL prediction system for EV batteries.

The EV battery multi variate time-series data such as cycle index, voltage, current and time-stamped operating metrics are collected. These raw data attained from on-board sensors or BMS, then preprocessed comprising of fundamental data cleansing steps like outlier deletion, missing value interpolation, and noise suppression by means of smoothing filters. The preprocessing approach ensures data reliability and integrity, which are instrumental for learning downstream. After preprocessing, the data is visualized using EDA, where statistical profiling, distribution mapping, and correlation heatmaps are created, reveals hidden patterns and associations between features. The visualized features, are normalized into specific range by min-max normalization. Then, data is split into training and testing sets, preserving chronological integrity for ensuring unbiased performance evaluation. Training set is employed to tune model parameters, and testing set is used to test generalization ability.

Hyperscale-cascaded TransformerNet architecture, a DL architecture with repeated transformer blocks structured in a cascaded hierarchy. Every transformer block consists of multi-head self-attention mechanisms, feed-forward layers, and positional encoding schemes that allow the model to learn long-range temporal dependencies and nonlinear degradation paths in battery dynamics.

III. PROPOSED SYSTEM MODELLING

A. Data Acquisition

The data acquisition phase kicks off the predictive pipeline by retrieving various operational parameters from EV battery systems. Some of the time-series data include cycle index, voltage, current, and other time-dependent features, which are recorded through embedded sensors in the battery management infrastructure. The acquisition layer is a specialized subsystem tasked with logging and forwarding these signals

to a centralized repository for data. It doesn't include the storage server itself, which belongs to downstream data analysis module.

New generation EV controllers are able to stream internal diagnostic signals e.g., charge/discharge rates, temperature profiles, and impedance metrics via socket-based communication protocols. The design enables data transmission in real-time or event-driven, depending on whether an external monitoring agent opens the connection. In this system, the data server controls the data acquisition process by deciding when to start or end data gathering and which battery modules or vehicle units to capture. This scalable and flexible communication configuration provides high-fidelity battery health data for ongoing preprocessing and predictive modelling.

B. Data Preprocessing

Data Cleaning

Smart meter datasets may have missing entries, noise, and anomalous readings owing to sensor faults or transmission errors. All missing values and detected outliers are replaced with the average energy use measured at the same hour on the previous day and the last good timestamp, maintaining temporal continuity and minimizing bias.

Exploratory Data Analysis

(a) Descriptive Statistics

Aggregate statistics like mean, median, standard deviation, and range are calculated to gain insight into central tendencies and variability in consumption patterns.

(b) Distribution Analysis

Histograms and density plots for visualizing spread and skewness of energy consumption over time assist in detecting non-Gaussian behavior or seasonal patterns.

(c) Correlation Heat maps

Pairwise correlation matrices are plotted to investigate the relationships between variables like hourly consumption, temperature, and day-of-week indicators. This supports feature selection and dimensionality reduction.

C. Feature Engineering (Min-Max Scaling)

To mitigate for the effects of scale differences in input data and improve prediction model performance, normalization is employed to map

raw features into a bounded range. This provides stable scaling and maintains the learning process steady.

Take an input sequence of energy consumption:

$$x = \{x_1, x_2, \dots, x_n\} \quad (1)$$

Where, (n) is the length of the sequence, each component (x_i) is normalized with the min-max scaling function:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

In the same fashion, for the output variables (y'_i) , the transformation is:

$$(y'_i) = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (3)$$

$x_{\min}, x_{\max}, y_{\min}, y_{\max}$ are the minimum and maximum limits of the respective variables. The normalized inputs (x'_i) and outputs (y'_i) obtained range between $[0, 1]$ so that they become compatible with activation functions and also facilitate better convergence during training.

D. Train and Test Split

The datasets are divided into two segments for model prediction analysis: 70% is designated for training the model, while the remaining 30% is reserved for testing its performance. The final model configuration is selected based on its accuracy achieved during the training phase.

E. Hyperscale Cascaded Transformer Net

Embedding Layer

In the novel Hyperscale-Cascaded TransformerNet approach to predict RUL of battery, the embedding layer is the first transformation phase that converts raw input sequences into a deep temporal modelling-friendly format. The input data is extracted using a sliding window strategy and organized as a batch of sequences of shape, $X \in \mathbb{R}^{B \times L \times F}$ where B is batch size, L is window length, and F is the number of battery health indicators.

The input is projected into a higher-dimensional representation $\mathbb{R}^{B \times L \times D}$, enabling the transformer encoder to capture subtle degradation patterns. Each input feature is converted to an embedding vector using a linear transformation:

$$E = XW_e \quad (4)$$

Where, $W_e \in \mathbb{R}^{F \times D}$ is a learnable embedding weight matrix. This operation maps each battery feature into a dense vector space that is more suitable for temporal learning.

Positional Encoding

Positional encoding introduces temporal order information into the input embeddings, allowing the model to differentiate between identical values occurring at different cycle indices. A sinusoidal positional encoding scheme is employed, defined as:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/D}}\right) \quad (5)$$

$$PE(pos, 2i+1) = \sin\left(\frac{pos}{10000^{2i/D}}\right) \quad (6)$$

Where pos denotes the time-step index, i is the embedding dimension index, and D is the embedding dimension. The positional encoding vectors are element-wise added to the input embeddings, ensuring the transformer retains temporal degradation ordering.

Transformer Block

The transformer encoder employs Multi-Head Self-Attention (MHSA) as given in fig.2, to capture long-range dependencies in battery degradation sequences.

Each embedded input is linearly projected into Query (Q), Key (K), and Value (V) matrices:

$$Q = ZW_Q, K = ZW_K, V = ZW_V \quad (7)$$

Where, $Z \in \mathbb{R}^{L \times D}$ is reshaped input tensor and W_Q, W_K, W_V are learnable parameters.

From fig.3, the Q, K, and V matrices are split into h attention heads:

$$Q, K, V \rightarrow \{Q_h, K_h, V_h\}_{h=1}^H \quad (8)$$

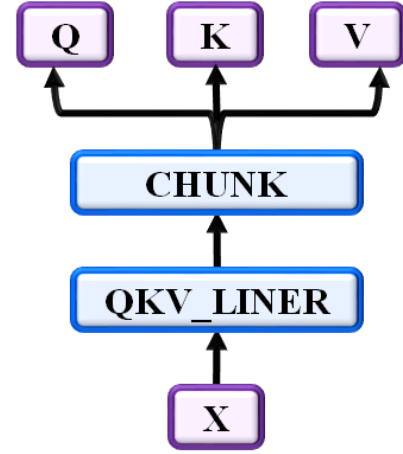


Fig. 2. Q, K, and V matrices generation.

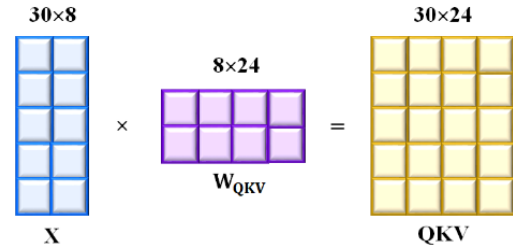


Fig. 3. The input data passes through a weight matrix.

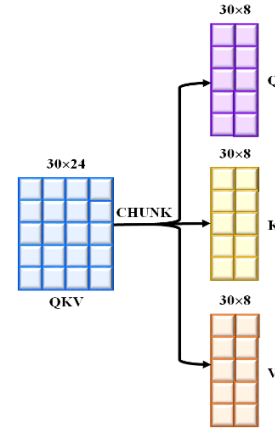


Fig. 4. Splitting of the Q, K, and V matrices.

Fig. 4, each head computes scaled dot-product attention:

$$Attention(Q_h, K_h, V_h) = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h \quad (9)$$

Where, d_k is the key dimensionality used to stabilize gradient propagation.

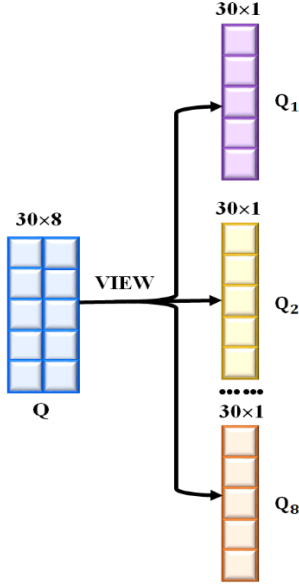


Fig. 5. Division of the attention head matrix.

The outputs of all heads are concatenated and linearly projected as displayed in fig.5:

$$MHSA(Z) = \text{Concat}(\text{head}_1, \dots, \text{head}_H) W_o \quad (10)$$

This process enables model to focus on several degradation patterns at once, enhancing its capacity to generalize across a wide range of battery health traces.

Multi-Layer Perceptron (MLP)

Following MHSA, a two-layer MLP refines the learned features using nonlinear transformations:

$$MLP(X) = \sigma(XW_1 + b_1)W_2 + b_2 \quad (11)$$

Where, W_1, W_2 are weight matrices, b_1, b_2 are bias vectors, and $\sigma(\cdot)$ denotes a nonlinear activation function.

This layer boosts the ability of the model to learn sophisticated degradation dynamics and nonlinear interactions between battery features.

Layer Normalization and Residual Connections

Layer Normalization (LN) is applied prior to both MHSA and MLP blocks, and residual connections are incorporated to preserve gradient flow:

$$Y = X + \text{Block}(\text{LN}(X)) \quad (12)$$

This design improves convergence stability and ensures sensitivity to battery health variations.

Feedforward Network Layer:

The feedforward network increases representational capacity while maintaining computational efficiency by using an expansion factor of two:

$$FFN(X) = \text{Dropout}(W_2 \text{ReLU}(W_1 X + b_1) b_2) \quad (13)$$

Here, (X) is the input feature matrix, and (W_1, W_2) represent weight matrices for the first and second linear layers, respectively. The bias vectors (b_1) and (b_2) are added to provide translation in the feature space. The first transformation $W_1 \in \mathbb{R}^{d_{\text{ff}} \times d_{\text{model}}}$ increases the input dimension (d_{model}) to an intermediate dimension $d_{\text{ff}} = 2 \times d_{\text{model}}$. Where the first linear layer expands the feature dimension and the second compresses it back to the original hidden size.

The Fig. 6 depicts the proposed hyper scale-Cascaded TransformerNet in predicting battery RUL.

Cascaded Transformer Block

The cascaded transformer block enhances RUL prediction by computing queries in the frequency domain while preserving keys and values in the temporal-spatial domain, enabling simultaneous modelling of long-term degradation trends and localized temporal dynamics. Consider the normalized battery health indicator is expressed as,

$$X \in \mathbb{R}^{L \times C} \quad (14)$$

Where, L represents temporal window length and C denotes number of feature channels.

Stage 1: Temporal-Spatial Attention Construction

In the first stage, the input tensor is projected into query, key, and value representations using point-wise and depth-wise convolutions:

$$Q_1 = \Phi_3\{\psi_1(X)\}, K_1 = \Phi_3\{\psi_1(X)\}, V_1 = \Phi_3\{\psi_1(X)\}, \quad (15)$$

Where, $\psi_1(\cdot)$ denotes 1×1 point-wise convolutions for channel mixing and $\Phi_3(\cdot)$ represents depth-wise convolution with kernel size 3 for local temporal feature extraction.

The attention mechanism proceeds as follows:

$$X' = \psi_1(\text{Attention}(Q_1, K_1, V_1)) \quad (16)$$

Stage 2: Frequency-Domain Feature Processor (FDFP)

In the second stage, only the query tensor is transformed into the frequency domain to capture global degradation characteristics, while keys and values remain in the temporal domain:

$$Q_2 = \mathcal{F}^{-1}(\mathcal{M} \odot \mathcal{F}(Q_1)) \quad (17)$$

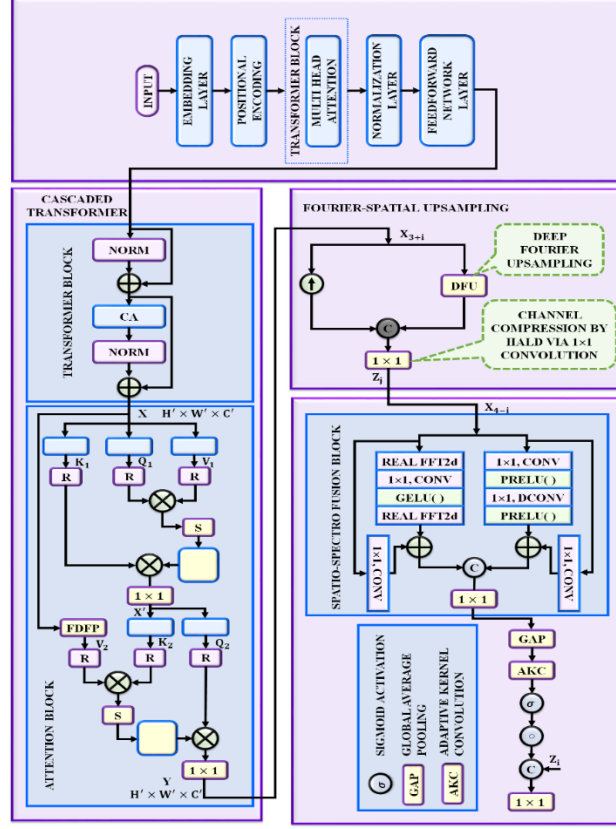


Fig. 6. Hyper scale-Cascaded TransformerNet.

Where, $\mathcal{F}(\cdot)$ and $\mathcal{F}^{(-1)}(\cdot)$ denote the 1D Fast Fourier Transform (FFT) and inverse FFT, respectively, and \mathcal{M} is a frequency mask emphasizing degradation-relevant bands.

Formal Definition of Frequency-Domain Operators:

The 1D Discrete Fourier Transform (DFT) of a temporal signal $x[n]$ of length L is defined as:

$$X[f] = \sum_{n=0}^{L-1} x[n] e^{-j2\pi fn/L}, \quad f = 0, \dots, L-1 \quad (18)$$

The Fast Fourier Transform (FFT), denoted by $\mathcal{F}(\cdot)$, is an efficient computational realization of the DFT.

The inverse DFT (IDFT), implemented via inverse FFT $\mathcal{F}^{-1}(\cdot)$, reconstructs the time-domain signal:

$$x[n] = \frac{1}{L} \sum_{f=0}^{L-1} X[f] e^{-j2\pi fn/L} \quad (19)$$

Frequency Masking and Degradation Bands

The operator \mathcal{M} denotes a frequency-domain mask that selectively emphasizes degradation-relevant spectral components:

$$\mathcal{M}(f) = \begin{cases} 1, & f \in \mathcal{F}_{deg} \\ 0, & otherwise \end{cases} \quad (20)$$

Here, \mathcal{F}_{deg} typically corresponds to:

Low-frequency bands: long-term aging and capacity fade

Mid-frequency bands: cyclic operational stress

High-frequency bands: noise and transient disturbances

This selective filtering ensures that the frequency-enhanced queries retain meaningful degradation information while suppressing irrelevant noise.

The frequency-enhanced queries are then used in attention computation:

$$X'' = \alpha \text{Attention}(Q_2, K_1, V_1) + X' \quad (21)$$

Here, α is a learnable scaling parameter that adjusts attention strength.

Spatio-Spectro Fusion-Based Attention for Temporal Feature Enhancement

To avoid propagating redundant temporal information, the Spatio-Spectro Fusion-Based Attention Block (SSFB) integrates spatial and spectral cues. The encoder feature E is processed as:

$$Y = \text{Conv}_{1D}([\text{SSFB}(E), D]) \quad (22)$$

SSFB operates through parallel spatial and frequency pathways:

$$\text{SSFB}(X) = \text{Concat}(\text{Spatial}(X), \text{Spectral}(X)) \quad (23)$$

Adaptive kernel sizing prevents over-smoothing:

$$k = \frac{\log_2(C)}{2} + 1 \quad (24)$$

Hybrid Fourier-Spatial Upsampling for RUL Feature Reconstruction

A hybrid upsampling strategy is used to reconstruct high-resolution degradation features:

Fourier-Domain Upsampling:

Zero-padding is performed on the frequency coefficients:

$$X_{up}(f) = \begin{cases} X(f), & f \leq f_{max} \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

then the inverse FFT is performed, allowing the recovery of smooth long-term degradation trends.

Pixel-Shuffle Spatial Upsampling:

The channel information is rearranged into a higher temporal resolution, thus the local transitions and cycle-level variations are better.

The hybrid method here is able to keep global aging patterns as well as the fine-grained temporal details, thus the accuracy and robustness of battery RUL prediction have been enhanced.

IV. RESULT AND DISCUSSION

This section outlines the main results through graphical results plotted in the form of Python-generated plots that give easy insight into the working of the system. Additionally, comparative analysis across datasets as well as classifier

systems is given, emphasizing the improved efficiency of proposed method with respect to state-of-the-art methods currently popular.

Table 1 Dataset Description.

Dataset Details	
Dataset name	Battery Remaining Useful Life (RUL)
Train Ratio	80%
Test Ratio	20%
No. of. Columns	15064
No. of. Rows	9
Data Source	Publicly available Kaggle dataset for battery remaining life prediction
Number of Batteries	The underlying data represent measurements from 14 individual lithium-ion batteries (NMC-LCO 18650 cells) that are cycled until failure.
Number of Cycles	Each battery is subjected to repeat charge-discharge cycles (typically on the order of hundreds to over a thousand cycles) until degradation to end-of-life, allowing RUL to be computed for each cycle.

Table.1 displays the proposed research utilizes Battery Remaining Useful Life (RUL) dataset, and data has been divided into training and testing sets with 80% and 20% ratio, fig.7 for ensuing model evaluation is robust. It contains measurement in the form of 9 rows and 15, 064 columns, representing data collected from 14 lithium ion batteries are cycled to repeat charge/dis-charge cycles until the batteries are completely degraded. This datasets serves as a solid foundation to model battery degradation and development of proposed predictive model for estimating battery lifespan in energy storage applications.

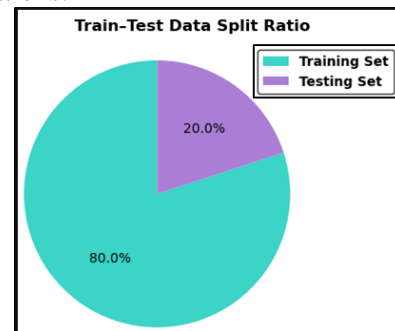


Fig.7. Train-Test data splitting.

Table 2 Feature Set and Target Variable for Battery RUL Modelling.

Cycle Index	Features
F1	Discharge Time (s)
F2	Time at 4.15 V (s)
F3	Time Constant Current (s)
F4	Decrement 3.6–3.4 V (s)
F5	Maximum Discharge Voltage (V)
F6	Minimum Charge Voltage (V)
F7	Charging Time (s)
Target Variable (RUL)	The Remaining Useful Life (RUL)

In table.2, dataset contains a set of features describing voltage and current behavior of lithium, ion batteries during their charge/discharge cycles. These are standard inputs in RUL modeling tasks.

Cycle Index is the sequential cycle number, and features like Discharge Time (F1), Time at 4.15 V (F2), and Time Constant Current (F3) represent the time intervals of the discharge and charging phases.

Decrement 3.63.4 V (F4) is the time for the voltage drop to occur within the critical range, and Maximum Discharge Voltage (F5) together with Minimum Charge Voltage (F6) give the voltage limits that show the battery's health.

Charging Time (F7) and Total Time are also measures of the energy transfer processes' duration.

These features, describe the electrochemical performance and decay of the batteries. The target variable is RUL, number of cycles through which the battery able to function until the end of life.

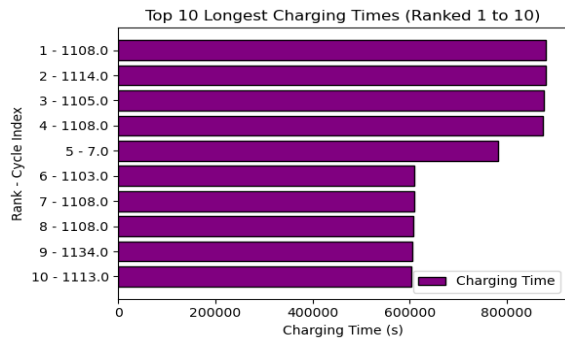


Fig. 8. Top longest charging times.

Fig. 8 presents the top 10 longest charging times across battery cycles in EV RUL dataset, indicating the top 10 battery cycles with the highest charging times in seconds. Each bar is assigned to a particular cycle index, and the x-axis is used for charging time and the y-axis for rank order from 1 to 10.

The visualization shows extreme variation in charging patterns between cycles, with many instances especially cycle indices 1108, 1114, 1105, and 1108 having prolonged charging times.

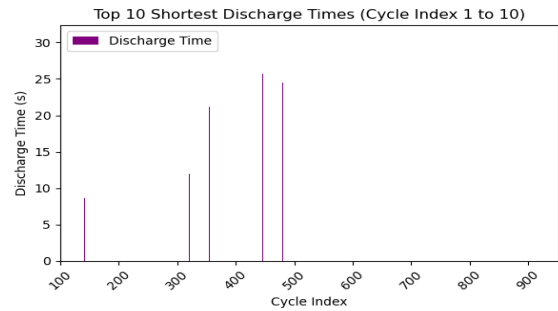


Fig. 9. Top shortest discharge times.

Fig. 9 represents the top 10 shortest discharge times through EV battery cycles, exhibiting the shortest discharge time's cluster around cycle indices 200, 400, and 500, where the minimum is seen nearest cycle index 200.

The shortened discharge times indicate lower energy retention, capacity degradation, or sudden voltage drops, all precursors to prompt aging.

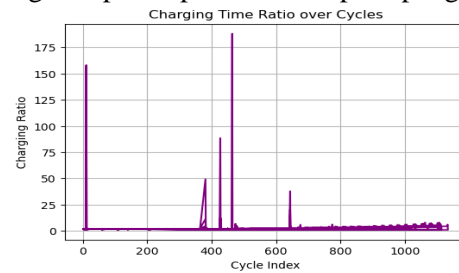


Fig. 10. Charging time ratio over cycles.

Fig. 10 displays the charging time ratio over cycles, it visually summarizes the temporal trend and anomalies in the charging pattern of a battery throughout its lifetime. The ratio of charging times is a sensitive measure of battery health, with well-behaved low values indicating normal use and spike-like occurrences like those at cycle indices 300 to 700 indicating possible degradation events, rising internal resistance, or heat stress. Detection of such outliers is very important for early fault detection and precise estimation of

RUL.

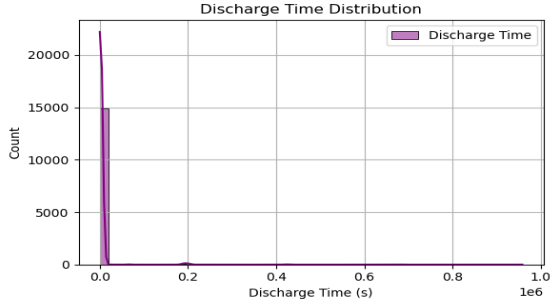


Fig. 11. Discharge time distribution.

Fig. 11 visualizes the temporal nature of battery discharge behavior over periods of operation. The histogram indicates that most discharge events take place in shorter time periods, with a dense concentration toward the initial discharge time. In the TransformerNet approach uses this discharge time data to learn latent representations of battery life, employing its multi-head attention mechanism to attend to both common short-duration events and the less common long-duration outliers. These temporal features are combined in the cascaded architecture to increase the robustness and granularity of RUL prediction.

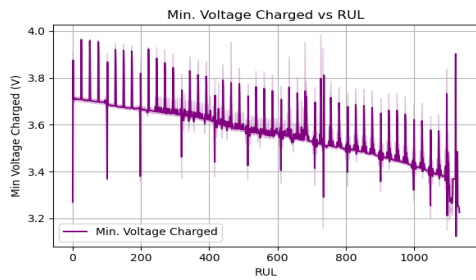


Fig. 12. Minimum voltage charged Vs RUL.

Fig. 12 illustrates the correlation between the minimum voltage charged and the remaining useful life of the battery, which exhibits an evident downward trend as RUL reduces. This reduction in minimum voltage is an indicator of progressive battery degradation, where lower voltage cut-offs indicate reduced capacity and rising internal resistance. The novel TransformerNet framework, establishes the predictive significance of voltage-based features, which are intrinsically electrochemically and thermally stress-sensitive and illustrates the capacity of the model to learn long-range dependencies and non-linear degradations through its cascaded attention mechanism.

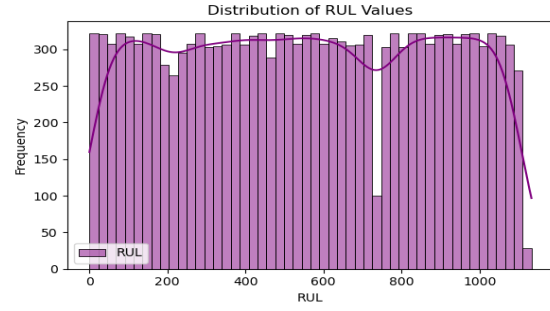


Fig. 13. Distribution of RUL values.

Fig.13 demonstrates the histogram overlaid with a line plot, showing the frequency distribution of RUL values in the dataset. The x-axis reflects the RUL in cycles, whereas the y-axis reflects the frequency of occurrence. The distribution is seemingly even, with a dip towards the 700-cycle point, potentially indicating a region of data sparsity or a phase shift in battery decline. Through both dense and sparse areas of the RUL spectrum, the model is able to better predict failure onset and learn to adjust its attention to key time periods.

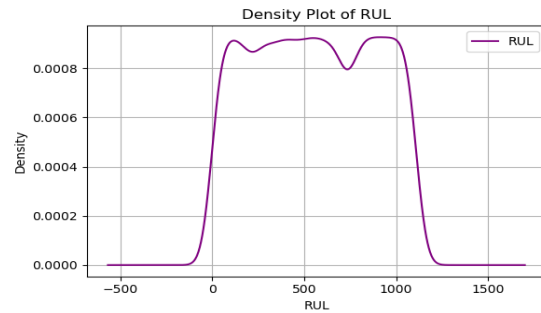


Fig. 14. Density plot of RUL.

Fig.14 reveals the probability density of RUL values throughout the dataset, providing a smooth, continuous flow of how often various RUL values are observed. The curve has a fairly flat top between 0 and 1000 cycles with minor dips at 250 and 750, and steep drop-offs at the extremes, which exhibits that most of the battery samples are in a middle operational lifespan range. The hyperscale-cascaded architecture is intended to deal with sophisticated, non-uniform distributions such as this one by dynamically modulating its attention on multiple temporal scales.

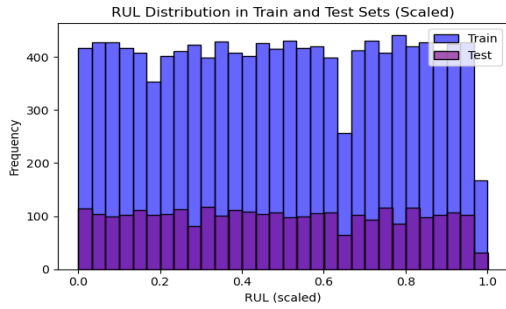


Fig. 16. RUL distribution in train and test sets.

Fig. 16 shows that the scaled distribution of RUL values over the training and test sets, providing information regarding the statistical balance and representational coverage of training and test data employed to train and test model. The hyperscale-cascaded structure, with its multi-head attention and hierarchical temporal modelling, is precisely tailored to learn from such rich input patterns.

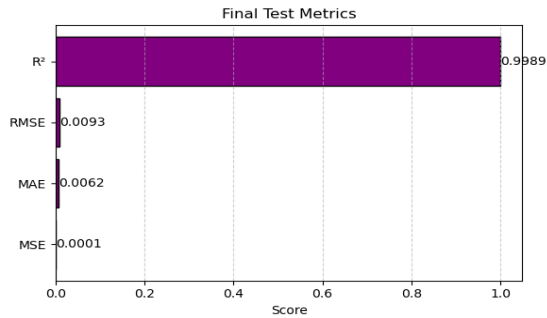


Fig. 17. Final test metrics.

Fig. 17 presents four important measure metrics R^2 , RMSE, MAE, and MSE which together evaluate the predictive strength and dependability of the model. The extremely high R^2 value of 0.9989 reveals that the model accounts for almost all the variance present in the RUL data, reflecting an excellent fit. In contrast, the very low error measures (RMSE: 0.0093, MAE: 0.0062, MSE: 0.0001) prove that the proposed hyperscale-cascaded TransformerNet model's predictions are accurate and regular, with negligible deviation from true values.

Fig. 18 represents the comparative line between actual RUL values (solid blue line) and predicted RUL values (dashed orange line) over a series of sample indices. The minimum variation in actual and predicted values throughout the sample range shows that the model generalizes and is robust even under situations of operating variability.

Fig.19 displays the ablation experiments comparing three transformers, based models:

Cascaded Transformer, HyperScale Transformer and HyperScale Cascaded Transformer.

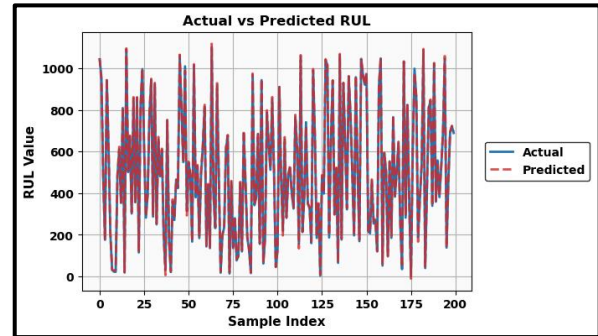


Fig. 18. Actual and Predicted values.

Each model is assessed by four performance metrics, including the R Score, MSE, MAE, and RMSE. From table.3, results indicate that the HyperScale Cascaded Transformer boasts the highest predictive accuracy with an R Score of 0.9993 with the lowest error values in all metrics (MSE: 0.0006, MAE: 0.0211, RMSE: 0.0245) results in better model convergence as well as lower prediction error in the forecasting task.

The evaluation of regression results between the HyperScale Cascaded Transformer and other current architectures (DNN and DenseNet) as presented in Fig 20 and Table 4 shows that the HyperScale Cascaded Transformer model provides an increased level of accuracy with a corresponding R^2 score and overall lower error based on all of the error metrics MSE of 0.0006, MAE of 0.0211, and RMSE of 0.0245 which points to the increased accuracy and stability of the predictive model.

Table.5 illustrates comparative analysis of different RUL prediction models in terms of their RMSE values. Out of the enumerated models, conventional methods like Artificial Neural Network-Random Forest-K-Nearest Neighbours-Gradient Boosting Decision Tree ANN-RF-KNN-GBDT [15], LSTM-Support Vector Machine (SVM) [16], Extreme Gradient Boosting-Light Gradient Boosting Machine, XGBoost-LightGBM [20] and 1D-CNN-BiLSTM [14] provides RMSE values of 11.007 to 0.8306, respectively, indicating low accuracy in achieving nonlinear degradation patterns. The proposed hyper scale-cascaded transformer net demonstrates a much lower RMSE of 0.0245, surpassing all other models. This huge reduction indicates model's capacity to catch both short-term fluctuations and long-term degradation

patterns via its multi-block transformer hierarchy and frequency-aware attention modules.

Fig. 21 visualizes the comparison of prediction error among prediction models employing MAE as the metric of evaluation. The developed model presents the lowest MAE of 0.0211, reflecting

better accuracy in predicting battery RUL, reflecting the stability and accuracy of the developed architecture, which renders it highly appropriate for real-time battery health monitoring and predictive maintenance in electric vehicle use.

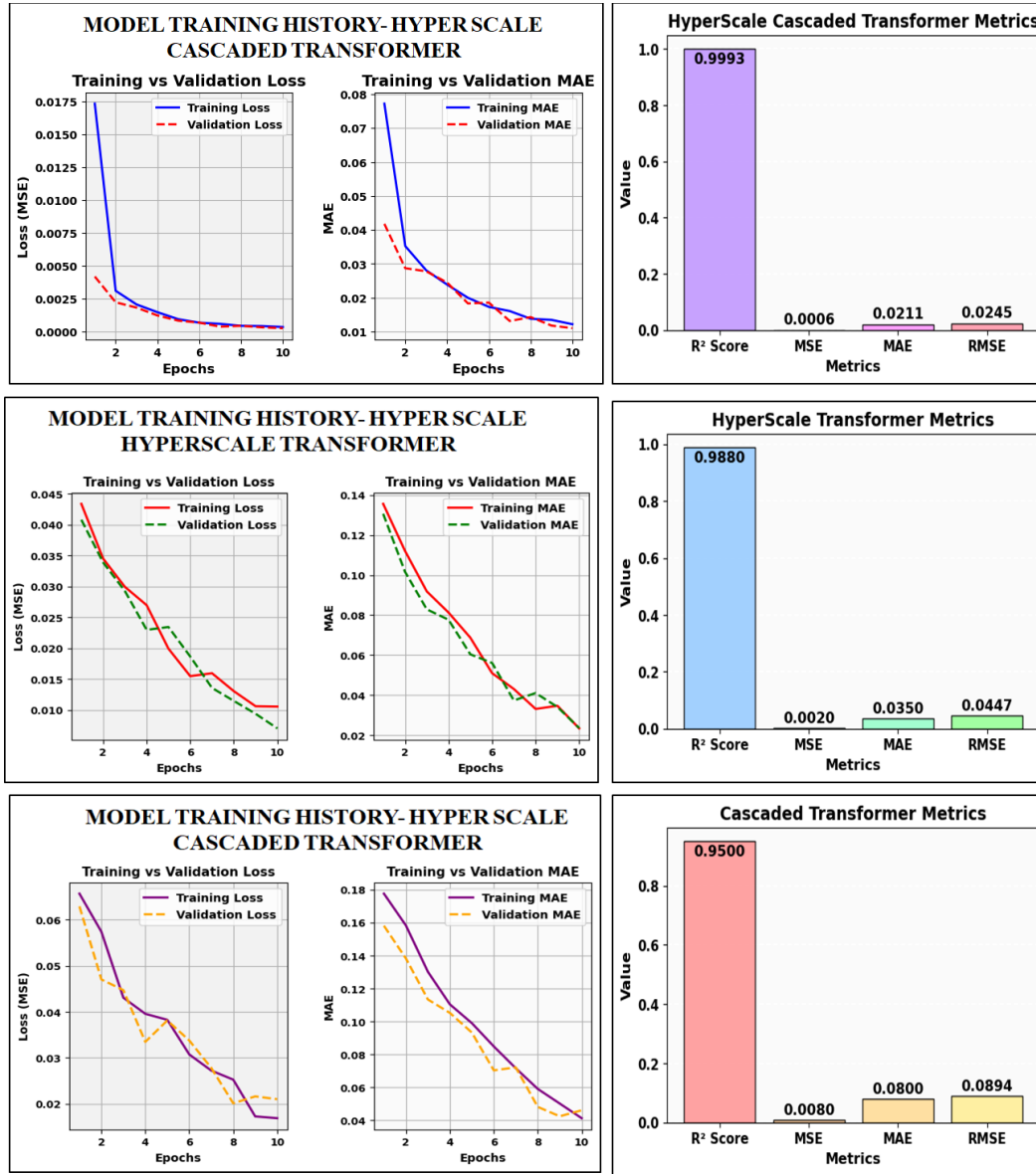


Fig.19. Ablation Studies.

Table.3 Ablation Study of Transformer-Based Architectures

Ablation Study				
Model	R ² -score	MSE	MAE	RMSE
Cascaded Transformer	0.95	0.008	0.08	0.0894
HyperScale Transformer	0.988	0.002	0.035	0.0447
HyperScale Cascaded Transformer	0.9993	0.0006	0.0211	0.0245

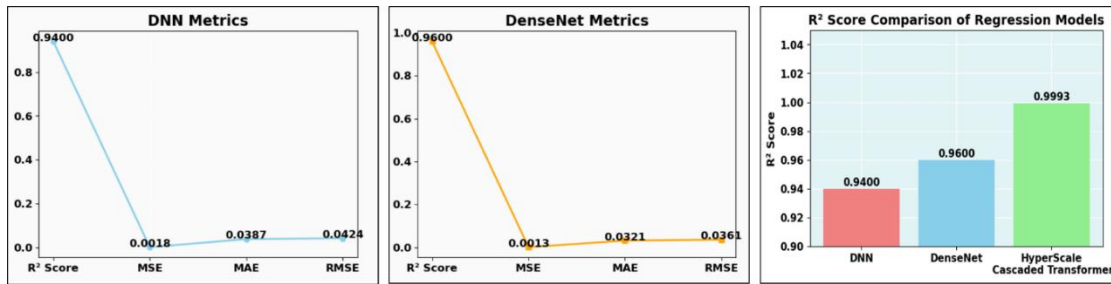


Fig.20. Comparison with existing model.

Table.4. Comparison of Regression Metrics with existing models.

Comparative analysis				
Model	R ² -score	MSE	MAE	RMSE
DNN	0.94	0.0018	0.0387	0.0424
DenseNet	0.96	0.0013	0.0321	0.0361
HyperScale Cascaded Transformer	0.9993	0.0006	0.0211	0.0245

Table 5 Comparison of RMSE

Prediction Model	RMSE
ANN-RF-KNN-GBDT [15]	11.007
LSTM-SVM [16]	6.9422
XGBoost-LightGBM [20]	0.94
1D-CNN-BiLSTM [14]	0.8306
Proposed	0.0245

Fig. 22 demonstrates a comparative study of MSE values for three predictive models LSTM-SVM) [16], RF [20], and the proposed technique, emphasizing the better accuracy of the recently proposed framework. The RF [20] model has the greater MSE at 1.67, showing high variance between estimated and actual values. The MLP [19] model with an MSE of 48.1941, presenting improved learning capability but still poor precision.

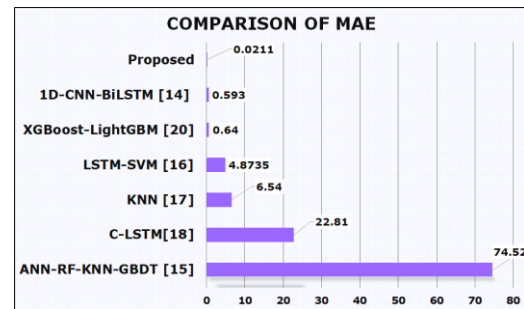


Fig. 21. Comparison of MAE.

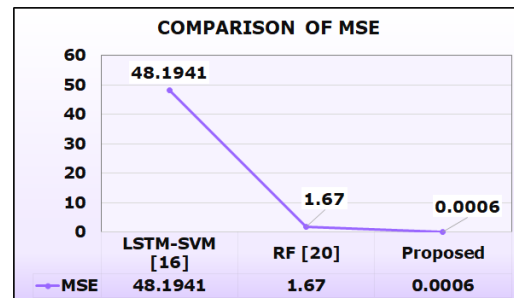


Fig. 22. Comparative analysis of MSE.

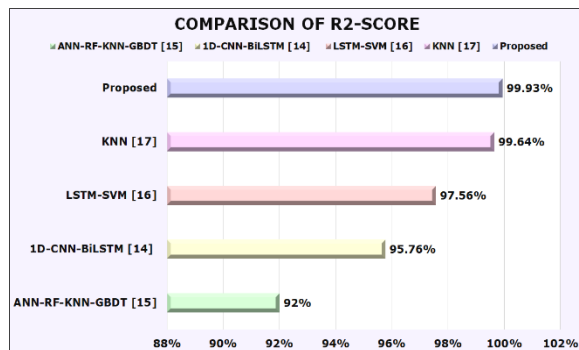


Fig. 23. Comparison of R2-Score.

Conversely, the proposed model has a very low MSE of 0.0006, which indicates confidence and forecasting accuracy of the proposed architecture, confirming its suitability for high-resolution battery RUL estimation and smart energy system prediction.

Fig. 23 compares the predictive ability of R²-score for different algorithms in predicting battery RUL. Out of the models compared proposed approach reflects better precision, with an R²-score of 99.93%, highlighting the superiority of

hyperscale-cascaded TransformerNet architecture in learning intricate temporal relations and degradation modes in EV battery data and yielding highly accurate and interpretable RUL predictions critical for predictive maintenance and energy optimisation.

V. CONCLUSION

This research presents a battery lifespan prediction system on the basis of a hyperscale-cascaded TransformerNet model that accurately captures the nonlinear dynamics of battery aging. The new model has very accurate predictions for differing battery capacity conditions and is extremely applicable for implementation in actual scenarios like data centres and grid-scale battery energy storage systems, where batteries are main energy storage or backup units. Model validation and performance testing were performed using Python, and the framework yielded a MAE of 0.0211, MSE of 0.0006, RMSE of 0.0245, and an R^2 -score of 0.9993. These performances identify the model's computational efficacy and adaptability for real-time battery health monitoring. In comparison to conventional ML methods, the Transformer-based method provides better accuracy and resilience, marking a milestone in EV battery management systems. To further improve predictive dependability, the next phase of research will investigate a wider variety of optimization techniques for predicting remaining battery capacity, incorporate more sophisticated feature selection algorithms, and examine domain-adaptive learning mechanisms to support various operating profiles and environmental scenarios.

REFERENCE

- [1] Celtek S.A., Kul S., Polat A.O., Zeinoddini-Meymand H., Shahnia, F. A Machine Learning-based Real-Time Remaining Useful Life Estimation and Fair Pricing Strategy for Electric Vehicle Battery Swapping Stations. *IEEE access*, 2025, vol. 13, pp. 62555 – 62566. <https://doi.org/10.1109/ACCESS.2025.3554682>.
- [2] Swain, Debabrata, Manish Kumar, Amro Nour, Kevin Patel, Ayush Bhatt, Biswaranjan Acharya, and Ali Bostani. "Remaining useful life predictor for EV batteries using machine learning." *IEEE Access* (2024).
- [3] Chu, Andrew, Anirudh Allam, Andrea Cordoba Arenas, Giorgio Rizzoni, and Simona Onori. "Stochastic capacity loss and remaining useful life models for lithium-ion batteries in plug-in hybrid electric vehicles." *Journal of Power Sources* 478 (2020): 228991.
- [4] Liu, Huiqiao, Qian Xiao, Yu Jin, Yunfei Mu, Jinhao Meng, Tianyu Zhang, Hongjie Jia, and Remus Teodorescu. "Improved LightGBM-based framework for electric vehicle lithium-ion battery remaining useful life prediction using multi health indicators." *Symmetry* 14, no. 8 (2022): 1584.
- [5] Geerts, David, Róbinson Medina, Wilfried van Sark, and Steven Wilkins. "Charge Scheduling of Electric Vehicle Fleets: Maximizing Battery Remaining Useful Life Using Machine Learning Models." *Batteries* 10, no. 2 (2024): 60.
- [6] Wu, Huawei, Congjin Ye, Yuanjin Zhang, Jingquan Nie, Yong Kuang, and Zhixiong Li. "Remaining useful life prediction of an igt module in electric vehicles statistical analysis." *Symmetry* 12, no. 8 (2020): 1325.
- [7] Pu, Ren, Shunli Wang, Xianpei Chen, Junhan Huang, Mingfang He, and Wen Cao. "A novel cuckoo search particle filtering strategy for the remaining useful life prediction of the lithium-ion batteries in hybrid electric vehicle." *International Journal of Energy Research* 46, no. 15 (2022): 21703-21715.
- [8] Gunapriya B., Karthik M., Abinaya I., Gandhi R.R., Vidhya H. Anti windup PI controller with tracking for motor drive system: Modelling, simulation and implementation in lab view based FPGA. *International Journal of Recent technology and Engineering*, 2020, vol. 8, no. 5. <https://doi.org/10.35940/ijrte.D9510.018520>
- [9] Wang, Xu, Jian Li, Ben-Chang Shia, Yi-Wei Kao, Chieh-Wen Ho, and Mingchih Chen. "A novel prediction process of the remaining useful life of electric vehicle battery using real-world data." *Processes* 9, no. 12 (2021): 2174.
- [10] Thirusenthil Kumaran P., Vinayagam A., Suganthi S.T., Veerasamy V., Inbamani A., Chandran J., Farade R.A. A voting approach of ensemble classifier for detection of power quality in islanded PV microgrid IETE Journal of Research, 2023, vol. 69, no. 10, pp. 7408-7424. <https://doi.org/10.1080/03772063.2022.2083706>.
- [11] Ansari Shaheer, Afida Ayob, Molla Shahadat Hossain Lipu, Aini Hussain, and Mohamad Hanif Md Saad. "Multi-channel profile based artificial neural network approach for remaining useful life prediction of electric vehicle lithium-ion batteries." *Energies* 14, no. 22 (2021): 7521.
- [12] Magadán Luis, Francisco J. Suárez, Juan C. Granda, Francisco J. delaCalle, and Daniel F. García. "A robust health prognostics

- technique for failure diagnosis and the remaining useful lifetime predictions of bearings in electric motors." *Applied sciences* 13, no. 4 (2023): 2220.
- [13] Li, Kai Ping Zhou, Yifan Lu, Xuebing Han, Xiangjun Li, and Yuejiu Zheng. "Battery life estimation based on cloud data for electric vehicles." *Journal of Power Sources* 468 (2020): 228192.
- [14] Gao, Dexin, Xin Liu, Zhenyu Zhu, and Qing Yang. "A hybrid CNN-BiLSTM approach for remaining useful life prediction of EVs lithium-Ion battery." *Measurement and Control* 56, no. 1-2 (2023): 371-383.
- [15] Huang, Weidong, Tong Zhou, Jiahuai Ma, and Xiaoyang Chen. "An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles." *Innovations in Applied Engineering and Technology* (2025): 1-12.
- [16] Sahdev, G., S. Agarwal, and P. Bhati. "Hybrid Learning Models for Remaining Useful Life Prediction of Lithium-Ion Batteries in Electric Vehicles." In *IOP Conference Series: Materials Science and Engineering*, vol. 1334, no. 1, p. 012003. IOP Publishing, 2025.
- [17] Sultan, Yara A., Abdelfattah A. Eladl, Mohamed A. Hassan, and Samah A. Gamel. "Enhancing electric vehicle battery lifespan: integrating active balancing and machine learning for precise RUL estimation." *Scientific Reports* 15, no. 1 (2025): 777.
- [18] Deng, Wujin, Yan Gao, Jianxue Chen, Aleksey Kudreyko, Carlo Cattani, Enrico Zio, and Wanqing Song. "Multi-fractal weibull adaptive model for the remaining useful life prediction of electric vehicle lithium batteries." *Entropy* 25, no. 4 (2023): 646.
- [19] Swain D., Kumar M., Nour, A., Patel K., Bhatt A., Acharya B., Bostani, A. Remaining useful life predictor for EV batteries using machine learning. *IEEE Access*, vol. 12, pp. 134418 – 134426, 2024. <https://doi.org/10.1109/ACCESS.2024.3461802>
- [20] Zhao Liang, Wei Yao, Yu Wang, and Jie Hu. "Machine learning-based method for remaining range prediction of electric vehicles." *IEEE Access* 8 (2020): 212423-212441.

Information about authors



CH. Siva Ganesh is pursuing his Ph.D. in Electrical Engineering at JNTUK University, A.P., India. His areas of interests are Power Electronics, FACTS, Power Quality. Email: ganesh6063@gmail.com
ORCID: 0000-0001-8625-7736



Dr. G. Satyanarayana, PhD degree in Electrical Engineering His areas of interests are Power Electronics, Renewable Energy Sources, Electrical Vehicles. Email: g.satyanarayana@giet.ac.in
ORCID: 0009-0003-1412-7205



Dontala Indu UG student. His areas of interests are Power Electronics, FACTS, Power Quality, Renewable Energy Sources, Electrical Vehicles. Email: dindu1052003@gmail.com
ORCID: 0009-0006-9309-8837



Chityala Veera Raghava Reddy UG student. His areas of interests are Power Electronics, FACTS, Electrical Vehicles. Email: raghavareddy6288@gmail.com
ORCID: 0009-0001-0711-8285



Gedela Lahari UG student. His areas of interests are Power Electronics, FACTS,. Email: gedelalahari70@gmail.com
ORCID: 0009-0006-4306-2465