

Received 26 June 2025, accepted 14 July 2025, date of publication 17 July 2025, date of current version 25 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3590001

RESEARCH ARTICLE

A Novel GRU-Augmented Time-Frequency Estimator for IGBT Remaining Useful Life Prediction

YUANKUANG LI¹, BAOHUA TAN^{1,2,3}, YINGJIE ZHANG⁴, (Member, IEEE),
SHUO WANG¹, BOLIN LIAN¹, YUAN LI¹, AND ZONGMING TAN⁵

¹Detroit Green Technology Institute, Hubei University of Technology, Wuhan, Hubei 430068, China

²School of Science (School of Chip Industry), Hubei University of Technology, Wuhan, Hubei 430068, China

³National "111 Research Center" Microelectronics and Integrated Circuits, Hubei University of Technology, Wuhan, Hubei 430068, China

⁴Department of Applied Physics, Faculty of Science, The Hong Kong Polytechnic University, Hong Kong, China

⁵Cogdel Cranleigh High School Wuhan, Wuhan, Hubei 432200, China

Corresponding authors: Baohua Tan (tbh@hbut.edu.cn) and Yingjie Zhang (yingjiey.j.zhang@connect.polyu.hk)

This work was supported in part by China University Industry-University Research Innovation Fund Project under Grant 2022BL052, in part by the Industry-University Cooperation and Education Program of the Ministry of Education under Grant 220900487303708, in part by the Major Project of Hubei Key Laboratory of Intelligent Transportation Technology and Device in Hubei Polytechnic University under Grant 2022XZ106, in part by the Talent Service Enterprise Project of Hubei Science and Technology Department under Grant 2023DJC119, and in part by the Program for College Students' Innovation and Entrepreneurship Training Projects under Grant S202410500082.

ABSTRACT Aging-induced failure of Insulated Gate Bipolar Transistors (IGBTs) significantly restricts the reliability of power electronic systems. Accurate and efficient prediction of IGBT Remaining Useful Life (RUL) is critical for proactive risk mitigation and ensuring system stability. Despite numerous existing aging models and data-driven methodologies, maintaining prediction robustness and accuracy under diverse and complex operational scenarios remains challenging. To overcome these limitations, we introduce a novel GRU-Augmented Time-Frequency Estimator (GATE) tailored for IGBT lifetime prediction. GATE utilizes an autoregressive time-series prediction framework trained via the Teacher Forcing strategy to recursively decode the electrical parameters indicative of the IGBT's aging state from rich time-frequency features. Experimental validations are performed using the square-wave power cycling dataset from the NASA Prognostics Data Repository. The results demonstrate that GATE significantly enhances prediction accuracy, reducing Mean Squared Error (MSE) to 0.0026 and Mean Absolute Error (MAE) to 0.045, representing improvements of 38.1% and 19.6%, respectively, compared to the leading baseline method. Moreover, recursive forecasting experiments show that GATE precisely predicts the remaining power cycles until the aging threshold (defined as a 15% increase in $V_{ce(on)}$) at various aging stages (10–60%). Ablation analyses further underline the critical contribution of the frequency-domain component. Collectively, these findings underscore GATE's capability to reliably decode IGBT RUL directly from historical operational data, bypassing intricate electrical or mechanical modeling, thereby offering a practically deployable and broadly generalizable solution for lifetime management in power electronic devices.

INDEX TERMS GRU, IGBT RUL prediction, teacher forcing, time-frequency fusion.

I. INTRODUCTION

Insulated Gate Bipolar Transistor (IGBT), characterized by low on-state voltage drop, excellent voltage-withstanding capability, high input impedance, and high modularity, have been widely applied in various power electronics fields,

The associate editor coordinating the review of this manuscript and approving it for publication was Jorge Esteban Rodas Benítez¹.

such as electric vehicles, aerospace, power transmission, and power management and control systems [1]. As critical components in power electronic converters, the reliability of IGBTs directly influences the stability of entire power systems. According to field failure reports, power electronic converters are common sources of failures in power systems, and such failures can have significant operational impacts [2]. For instance, in wind power plants, faults in power electronic converters account for 13% of total failures, leading to approximately 18% of downtime [3]. Similarly, in photovoltaic (PV) systems, inverter faults constitute around 55% of all failures, predominantly caused by failures in power electronic converters [4]. Statistics further show that IGBTs constitute about 42% of all power devices used [5], highlighting their critical role in system reliability. Consequently, predicting the IGBT RUL has become a prominent research topic.

The reliability of IGBTs is closely related to their operating conditions. Temperature fluctuations, arising from various factors, cause packaging aging, ultimately affecting the device's lifespan [6]. Given the high power density and frequency characteristics of IGBTs, they are extensively used in high-voltage, high-current control, and high-frequency switching applications. This inevitably exposes IGBTs to tens of thousands or even millions of thermal shocks throughout their operational lifecycle. However, before complete failure occurs, devices may still function normally for a certain period. Therefore, accurately predicting the RUL of IGBTs before complete failure allows timely maintenance and replacement, effectively preventing system faults.

To evaluate the aging condition of IGBTs and timely detect defective devices, researchers worldwide have extensively investigated methods for identifying and extracting fault-characteristic parameters. Some studies have discovered significant correlations between junction temperature and IGBT lifespan, leading to junction temperature-based RUL prediction methods [7]. However, directly measuring junction temperature poses considerable challenges, necessitating estimation through electrothermal parameters [5]. These parameters are influenced not only by junction temperature but also by chip degradation and packaging deformation, complicating the RUL prediction task based on junction temperature estimation. Some researchers proposed monitoring chip temperature using fiber-optic temperature sensors [8],

effectively bypassing complex modeling and computations but requiring costly dedicated thermal imaging equipment, limiting practical deployment.

Existing studies suggest that electrical parameters such as I_C , V_{GE} , T_{on} , T_{off} , and T_j considering measurement ease, accuracy, sensitivity, heritability, linearity, and calibration, are more suitable as early aging indicators of IGBTs [9]. Among these, the IGBT collector-emitter on-state voltage drop, $V_{ce(on)}$, is considered the optimal indicator for predicting IGBT RUL. During actual operation, voltage and current are not constant; fluctuating electrical stress leads to persistent temperature variations, causing cracks or voids in the solder layer, the primary heat dissipation channel of the IGBT. This reduces the thermal conduction area between the solder layer and the IGBT chip, increases thermal resistance, and consequently raises junction temperature [9]. Additionally, $V_{ce(on)}$ is a positive temperature-sensitive electrical parameter; thus, as junction temperature increases, $V_{ce(on)}$ correspondingly rises [10].

The predominant failure mechanism in IGBTs is bond wire lift-off, accounting for about 70% of IGBT module failures [11]. During operation, differences in coefficients of thermal expansion (CTE) among various internal layers cause non-uniform thermal and mechanical stresses under high voltage, high current, and high-frequency conditions. These stresses similarly induce cracks or voids in the solder layer. However, shear stresses also occur at the bond wire and chip interface during thermal cycling, leading to thermomechanical fatigue. Repeated thermal cycles progressively reduce the contact area between bond wires and the chip, increasing current density and stress accumulation until eventual bond wire lift-off. Concurrently, on-state resistance increases, continuously raising the $V_{ce(on)}$ until complete failure.

Clearly, the collector-emitter on-state voltage, $V_{ce(on)}$, can accurately reflect the RUL condition of IGBTs. Due to its ease of measurement, $V_{ce(on)}$ is adopted as the primary electrical parameter for RUL prediction. In engineering applications, complete device failure is typically not used as the failure standard; rather, an IGBT is deemed failed when its reliability degrades to a certain threshold. Some researchers define failure as a 20% increase in $V_{ce(on)}$ compared to its initial value [12], while others propose a 10% increase [5], [13]. Considering these perspectives, this paper selects an intermediate threshold of 15%.

Clearly, the on-state voltage drop ($V_{ce(on)}$) can accurately reflect the RUL state of IGBTs. Due to the convenience of its measurement, $V_{ce(on)}$ has become the primary electrical parameter employed for RUL prediction. In practical engineering applications, devices are typically not considered failed only upon complete functional loss; rather, failure is defined when IGBT reliability declines below a specific threshold. Some researchers define failure as a 20% increase in $V_{ce(on)}$ relative to its initial value [12], while others propose a 10% increment as the threshold [13]. Several scholars have also regarded increments within the range of 10% to 20% as indicative of degradation or failure.

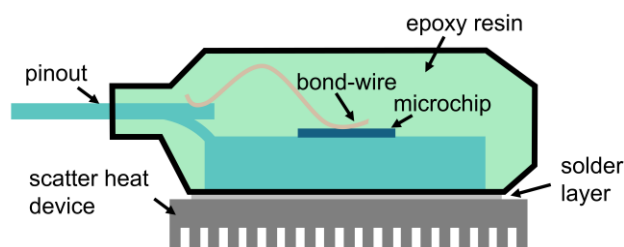


FIGURE 1. Single-tube IGBT exploded view.

Considering these research findings and the inherent sensitivity of $V_{ce(on)}$ to temperature and aging states, this study selects a 15% increment as the degradation threshold. This value, situated midway within the range recommended by the literature, avoids both premature failure determination—which could lead to prediction biases—and excessively high thresholds, which might neglect critical degradation stages, thereby offering strong representativeness and rationality.

Most traditional studies still rely heavily on empirical formulas or physics-based models to estimate RUL. However, these approaches generally assume idealized stress scenarios and struggle to accommodate complex load fluctuations and nonlinear degradation characteristics encountered in practical operational conditions. Additionally, traditional methods such as those based on Weibull distribution regression [14] are usually applicable only to specific failure modes, thereby limiting their generalizability and accuracy. In recent years, data-driven methods have gained widespread application in RUL prediction tasks, particularly models based on Recurrent Neural Networks (RNN), which demonstrate superior performance in capturing time dependencies within degradation sequences. Specifically, Long Short-Term Memory (LSTM) networks have been utilized to model IGBT degradation behavior [15], significantly enhancing prediction accuracy. Nevertheless, RNN-based models commonly suffer from error accumulation and insufficient prediction coherence in recursive multi-step predictions, posing challenges for industrial deployments, which require stringent model stability and operational efficiency.

Addressing these issues, this paper proposes a lightweight Gated Recurrent Unit (GRU)-based prediction framework. By integrating frequency-domain spectral amplitude features and thoroughly examining the Teacher Forcing training mechanism, the proposed approach enhances the model's capability to represent nonlinear degradation trends and maintains stability in multi-step predictions. Moreover, the method achieves high predictive accuracy and favorable deployment adaptability while maintaining low computational complexity, making it suited for embedded industrial systems.

The main contributions of this work are as follows:

1. We propose a lightweight GRU-based GATE model that integrates time-domain and FFT-derived frequency-domain features to improve RUL prediction accuracy and effectively characterize nonlinear degradation patterns.
2. A monotonicity constraint is introduced based on the physical properties of $V_{ce(on)}$, ensuring physically consistent and stable predictions throughout the device aging process.
3. Various Teacher Forcing strategies are systematically explored, and an optimal configuration is identified to alleviate error accumulation and enhance stability in multi-step forecasting.
4. Comprehensive experiments—including ablation studies, comparisons with Transformer and

physics-informed models, and deployment evaluations—validate the model's predictive performance and suitability for industrial applications.

II. RELATED WORK

A. IGBT RUL PREDICTION BASED ON AGING MODELS

As the lifespan degradation of IGBT modules results primarily from material aging and deformation—processes exhibiting certain predictability—several empirical mathematical and numerical models have been proposed to assess the RUL of power electronic devices. Among these, the LESIT model is particularly prominent. It was developed by integrating the Coffin-Manson fatigue law with the Arrhenius temperature acceleration model, enhancing predictive accuracy and applicability [16].

The basic form of the Coffin-Manson equation is given by:

$$N_f = C \cdot (\Delta \varepsilon)^m \quad (1)$$

In the equation: N_f represents the number of cycles before failure; C is a material constant that depends on the specific material and operating conditions; $\Delta \varepsilon$ denotes the strain range, representing the difference in strain during cyclic loading (in power electronic devices, $\Delta \varepsilon$ is rewritten as the junction temperature swing ΔT_J); m is the Coffin-Manson exponent, typically a negative value, indicating the sensitivity of fatigue life to the strain range.

The fundamental form of the Arrhenius equation is as follows:

$$k = A \cdot e^{-\frac{E_a}{k_B T}} \quad (2)$$

In the equation: k represents the reaction rate constant; A is the pre-exponential factor, which depends on the specific reaction mechanism; E_a denotes the activation energy; k_B is the Boltzmann constant; T represents the absolute temperature.

The core equation of the LESIT model for IGBT RUL prediction is as follows:

$$N_f = K \cdot (\Delta T_J)^{\beta_1} \cdot e^{\frac{\beta_2}{T_J + 273}} \cdot t_{on}^{\beta_3} \cdot I^{\beta_4} \cdot V^{\beta_5} \cdot D^{\beta_6} \quad (3)$$

In the equation: N_f represents the number of cycles before failure; ΔT_J denotes the junction temperature swing; I represents the current through each bond wire; V is the blocking voltage of the chip; D refers to the diameter of the bonding wire.

The current research indicates that the LESIT model serves as a well-established reference for predicting the IGBT RUL, providing satisfactory results. However, the LESIT equation is essentially a static lifetime model; it calculates the total lifespan of the device only after the actual failure has occurred, thus serving merely as an empirical estimate rather than dynamically predicting the exact time of failure in advance.

B. IGBT RUL PREDICTION BASED ON MACHINE LEARNING

Compared to traditional empirical mathematical models, deep learning methods possess the capability of dynamic prediction based on available data, enabling more effective predictions of the RUL of IGBT devices. However, despite numerous advancements in recent years, existing research still significantly lacks effectiveness in capturing the complex temporal-frequency degradation characteristics of IGBT aging processes. For instance, Tian et al. [17] proposed an adaptive boosting method based on a multi-class neural network, which improved classification accuracy but did not incorporate frequency-domain degradation information. Ali et al. [18] utilized Gaussian Process Regression (GPR) combined with Bayesian inference for RUL estimation, offering advantages over traditional Maximum Likelihood Estimation methods in uncertainty modeling, yet exhibited limited scalability with large-scale datasets. He et al. [19] compared the performance of various machine learning models, such as Backpropagation Neural Networks, Random Forests, and Extreme Learning Machines, on the NASA IGBT aging dataset, highlighting performance disparities but lacking comprehensive exploration in feature selection and degradation mechanism modeling. Wang et al. [20] further integrated Particle Swarm Optimization (PSO) with Random Forest models, combining time-domain and frequency-domain features for state prediction, significantly enhancing prediction accuracy through improved feature fusion and model optimization.

Despite these extensive explorations, a unified deep learning framework capable of jointly modeling temporal degradation trends and frequency-domain features remains lacking. To bridge this research gap, this paper proposes a hybrid deep learning model.

C. TEACHER FORCING STRATEGY

In traditional autoregressive models, predictions at each time step depend on outputs from the previous step, potentially causing slow convergence and error accumulation during sequence generation. To address this, the Teacher Forcing strategy always uses the true target value as the next input during training [21], rather than relying on the model's own predictions. This provides the model with accurate context information, accelerating convergence and mitigating gradient propagation difficulties.

However, this method faces distribution mismatch issues during testing. Since the true target values are unavailable at inference, models must rely solely on their own predictions, leading to "exposure bias," where generated results gradually deviate from true sequences. To mitigate this issue, researchers introduced the Scheduled Sampling strategy, which controls the balance between using true targets and model predictions during training through a probability ε . Initially, ε is high, making the model rely primarily on true targets; as training progresses, ε gradually decreases, enabling the model to use its own predictions

and better adapt to inference scenarios [22]. Alternatively, Professor Forcing introduces a discriminator to compare RNN behaviors under teacher forcing (using real data) and free-running modes (using generated data). Through adversarial training, Professor Forcing ensures consistency between hidden states and output distributions under both conditions [23].

III. GRU-AUGMENTED TIME-FREQUENCY ESTIMATOR FOR IGBT REMAINING LIFETIME PREDICTION (GATE)

A. PROBLEM DEFINITION

The IGBT RUL prediction task is a time-series forecasting problem. To accurately assess device health and remaining lifetime, this study defines RUL prediction as a univariate time-series forecasting problem based on $V_{ce(on)}$, which serves as a degradation indicator. Given a historical sequence $\{x_1, x_2, \dots, x_n\}$, the goal is to estimate the future trend of $V_{ce(on)}$ over the next time steps to support device lifetime management and failure prevention.

Assuming that degradation data from D IGBT devices is collected under continuous thermal cycling tests, the consolidated degradation curves can be represented as:

$$(x_1^i, x_2^i, \dots, x_{N_i}^i), \quad i = 1, 2, \dots, D \quad (4)$$

where x_n^i represents the $V_{ce(on)}$ value measured for the i -th IGBT device at the n -th thermal cycling round, and D denotes the total number of devices.

The model employs a GRU network for time-domain modeling of the $V_{ce(on)}$ sequence while integrating frequency-domain features extracted via Fast Fourier Transform (FFT). Mathematically, for each device sequence with a historical prefix $\{x_1^i, x_2^i, \dots, x_n^i\}$, the sequence is first processed through the GRU network to extract hidden representations that reflect aging evolution. Then, FFT is applied to the same prefix sequence, and the amplitude spectra of key frequency components are selected as covariates, which are then processed through an MLP to obtain a frequency-domain embedding. The final hidden state from the GRU and the MLP output are fused, and a fully connected layer is used to predict the $V_{ce(on)}$ value for the next thermal cycle x_{n+1}^i . Abstracting this process as a function F , the prediction function can be expressed as:

$$x_{n+1}^i = F(\text{GRU}(x_{1:n}^i), \text{MLP}(\text{FFT}(x_{1:n}^i))) \quad (5)$$

where $\text{FFT}(x_{1:n}^i)$ represents the frequency-domain information of the sequence at the given cycle.

The above model enables accurate prediction of the next $V_{ce(on)}$ value; however, in many cases, predictions over a longer time horizon are required. To achieve this, a recursive iteration approach is used for long-term forecasting.

The core of recursive prediction is to append the current predicted value $\hat{x}_{n+1}$ to the end of the historical sequence and use the updated sequence as input for the next step. Through this rolling iteration process, future values for multiple time steps can be continuously inferred.

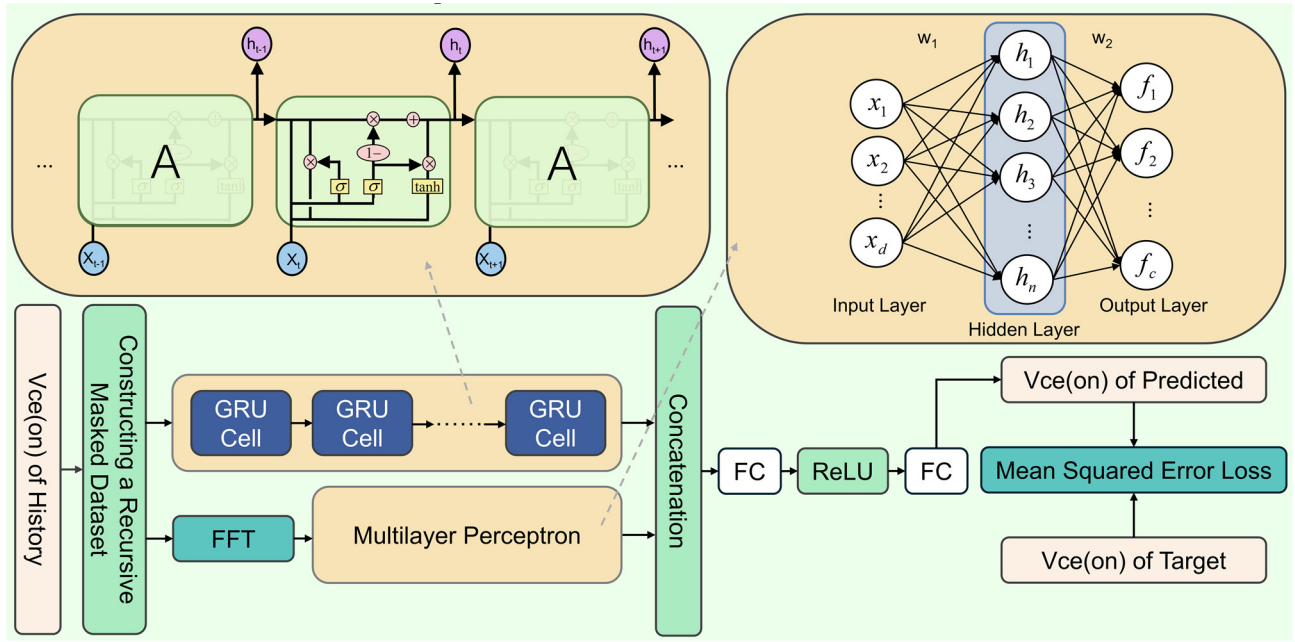


FIGURE 2. GATE model framework.

This process can be viewed as an explicit autoregressive modeling, expressed as:

$$x_{n+m}^i = F(\hat{x}_{n+m-1}^i, \hat{x}_{n+m-2}^i, \dots; \theta) \quad m = 1, 2, \dots, T \quad (6)$$

where θ represents the learnable model parameters, which are the trainable network weights that need to be optimized through backpropagation and gradient descent during training.

B. MODEL OVERVIEW

The proposed GATE model in this study aims to comprehensively mine both temporal and frequency-domain features hidden within the one-dimensional historical $V_{ce(on)}$ data, thus enabling precise regression predictions of IGBT module RUL. Figure 2 illustrates the overall architecture of the proposed model, which primarily consists of two feature extraction branches. The first is a time-domain branch, where raw historical data is input into a GRU network to extract temporal features. The recurrent structure of GRU effectively captures temporal dependencies, generating high-dimensional hidden states. The second branch transforms the raw historical data into frequency-domain signals using FFT, extracting spectral amplitude features and subsequently feeding these into an MLP for nonlinear mapping and feature extraction. After extracting features from both branches, the outputs are concatenated and integrated, followed by linear transformations and nonlinear activation through fully connected layers, ultimately providing the predicted RUL of the IGBT modules.

C. GATED RECURRENT UNIT (GRU)

Traditional recurrent neural networks (RNNs) often suffer from gradient vanishing or exploding problems during

training, making it difficult for the model to remember long-term and short-term information [24]. To address this issue, Long Short-Term Memory (LSTM) networks introduce gate mechanisms, effectively managing information retention and forgetting, thereby alleviating gradient-related problems to some extent [25], [26]. The GRU model, an optimized variant derived from LSTM, simplifies the structure without introducing additional parameters by merging the input and forget gates into a single update gate and combining cell states and hidden states [27]. GRU models maintain accuracy while having fewer training parameters and faster convergence than LSTM models [28].

The reset gate determines whether to forget the previous step's information by combining the hidden state from the previous time step h_{t-1} and the current input x_t . The result is passed through a tanh activation function to generate the candidate hidden state h'_t .

The update gate controls the extent to which information is retained. It produces z_t using a sigmoid activation function, where z_t ranges between 0 and 1. When z_t is closer to 1, more information is preserved; when it is closer to 0, less information is retained.

The computation formulas are as follows:

$$r^t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (7)$$

$$h'_t = \tanh(W_{h'} \cdot [x_t, r^t \odot h_{t-1}]) \quad (8)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t \quad (10)$$

In the equation: x_t and h_t represent the current input and the new hidden state of the GRU unit at time step t , respectively; h_{t-1} is the hidden state from the previous time step;

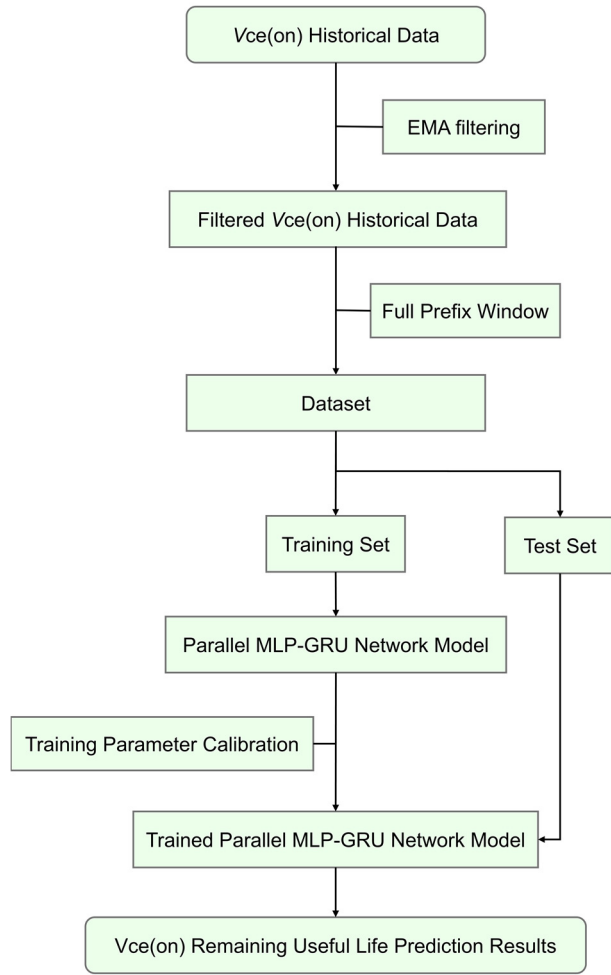


FIGURE 3. Model training and prediction flowchart.

$[ht-1, xt]$ denotes the concatenation of the previous hidden state and the current input along the feature dimension; W_γ , $W_{ht'}$ and W_z are the weight matrices corresponding to the reset gate, candidate hidden state, and update gate, respectively; $\sigma(\cdot)$ represents the sigmoid activation function; h'_t represents the sigmoid activation function; $\cdot$ denotes matrix multiplication; $\odot$ represents element-wise multiplication.

In time-series prediction tasks, GRU is particularly suitable due to its lightweight characteristics, making it advantageous for scenarios involving smaller datasets or real-time requirements. In contrast, LSTM networks typically involve more parameters and memory usage, and a complete Transformer further increases parameter scale significantly. Thus, this paper integrates the Teacher Forcing training mechanism with the GRU's temporal processing capability to construct an efficient and lightweight prediction model.

D. TIME-FREQUENCY FUSION

In time-series modeling tasks, particularly for modeling and predicting the degradation process of IGBT, key signals such as $V_{ce(on)}$ not only exhibit monotonic trends evolving over

time but also contain significant periodicity and spectral characteristics. Existing research indicates that under prolonged thermal-electrical cyclic stress, the spectral response of $V_{ce(on)}$ changes markedly with aging: the energy within its high-frequency components (1–10 MHz) gradually attenuates, accompanied by a consistent decrease in the amplitude of dominant spectral peaks [29]. Such frequency-domain evolution patterns reflect physical degradation processes within the device's internal structures, such as bond wire cracks and solder joint aging, offering enhanced sensitivity and interpretability for degradation monitoring and potentially serving as early indicators of device failure.

Traditional models based solely on time-domain features, such as GRU, exhibit limitations in capturing these frequency-domain degradation signals. To address this gap, this study extends the GRU's capability in extracting temporal characteristics by introducing an additional frequency-domain modeling branch. Historical sequence signals are mapped from the time domain to the frequency domain using FFT, extracting spectral amplitude and phase information. For unified modeling, the resulting spectral vectors are standardized to fixed lengths via truncation or zero-padding and subsequently processed by a Multi-Layer Perceptron (MLP) to explore nonlinear and higher-order relationships.

Ultimately, the frequency-domain features are combined with the temporal hidden states generated by the GRU within a feature fusion module, facilitating integrated modeling for subsequent RUL prediction. This dual-branch architecture not only preserves the capability to model dynamic temporal evolutions but also enhances sensitivity to latent periodicity and spectral degradation patterns, thereby significantly improving the model's adaptability, robustness, and generalization in capturing both short-term fluctuations and long-term trends in the IGBT degradation process.

IV. EXPERIMENTAL SETUP AND RESULTS

A. EXPERIMENTAL SETUP

This study utilizes the accelerated degradation dataset of $V_{ce(on)}$ collected by NASA Ames Research Center [30], which has gained significant recognition and extensive citation in the field of power device lifetime modeling and prediction. The data were acquired by NASA researchers through classical thermal cycling experiments, wherein IGBT devices were repeatedly heated and cooled to simulate thermal stress conditions commonly encountered in practical industrial environments [31]. The experiment involved six IGBT devices of the IRG4BC30KD type, each installed on a customized aging test platform. Each device underwent periodic excitation controlled by Pulse Width Modulation signals with a frequency of 10 kHz and a duty cycle of 40%. The ambient temperature varied precisely between 20°C and 100°C, ensuring accurately controlled thermal stress conditions. The $V_{ce(on)}$ signal was continuously recorded with high temporal resolution throughout the devices' complete degradation

process, comprising over 100,000 data samples. The data acquisition system consisted of high-precision oscilloscopes and data acquisition cards, achieving well-controlled noise levels without significant external interference.

Considering that $V_{ce(on)}$ is influenced not only by aging but also by junction temperature, this study employs a linear temperature-dependent model of $V_{ce(on)}$ to mitigate the effect of junction temperature within the dataset [32]. The equation is given as follows:

$$V_{CE} = V_0 + k_T \cdot (T_j - T_0) \quad (11)$$

In this equation, $V_{ce(on)}$ denotes the collector-emitter on-state voltage measured at the current junction temperature T_j , while V_0 represents the reference on-state voltage drop at a baseline temperature T_0 (typically 25°C). K_T is the temperature coefficient, indicating the incremental change in the on-state voltage per 1°C increase in temperature.

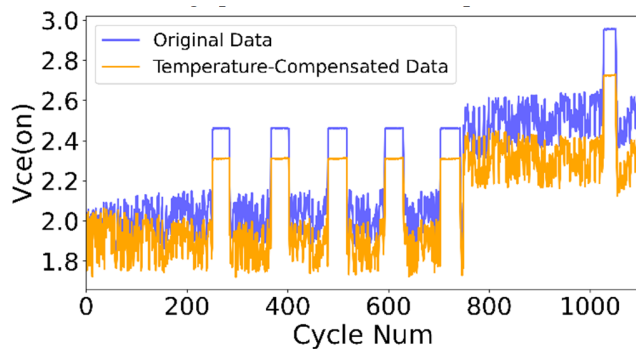


FIGURE 4. Temperature compensation effect on $V_{ce(on)}$.

This linear model has been extensively utilized in temperature compensation and device health monitoring applications. Based on this model, this study implements temperature compensation on the measured on-state voltage data, effectively eliminating disturbances caused by temperature fluctuations. Consequently, the degradation in electrical performance due to device aging or damage can be more accurately identified. Figure 4 shows the compensated image.

To effectively suppress random noise, data smoothing was conducted using Exponential Moving Average (EMA). Unlike the fixed-window moving average, EMA assigns different weights to recent and historical data points, maintaining signal trends while dynamically smoothing the data. The EMA calculation formula is as follows:

$$EMA_t = \alpha \times x_t + (1 - \alpha) \times EMA_{t-1} \quad (12)$$

In the equation: x_t represents the data value at time step t ; EMA_t denotes the exponentially weighted moving average at time step t ; α is the smoothing coefficient, which ranges between 0 and 1.

In this study, α is set to 0.3, meaning that in each computation step, 30% of the weight is assigned to the new data, while 70% of the weight comes from the previous filtered value. After applying the EMA filter, the data is normalized using

the Min-Max normalization method, which maps the results into the [0,1] range. The transformation function is given as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

In the equation: $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of the sample data.

Since the IGBT lifespan follows a monotonic trend over time, this study enforces an increasing constraint on $V_{ce(on)}$ to ensure a strictly rising trend throughout the aging process.

$$\hat{y}_t = \max\{\tilde{y}_t, \hat{y}_{t-1} + \varepsilon\} \quad (14)$$

In the equation: $\tilde{y}_t$ represents the raw prediction of the model at time step t ; $\hat{y}_t$ denotes the adjusted prediction after applying the increasing constraint; $\hat{y}_{t-1}$ is the prediction from the previous time step; ε is a small positive constant to ensure strict monotonicity.

This study evaluates the difference between the model's predictions and the actual values using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The MSE metric amplifies larger errors due to squaring, making it more sensitive to significant prediction deviations.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (15)$$

In the equation: n represents the total number of samples; y_i denotes the actual value of the i -th sample; $\hat{y}_i$ denotes the predicted value of the i -th sample.

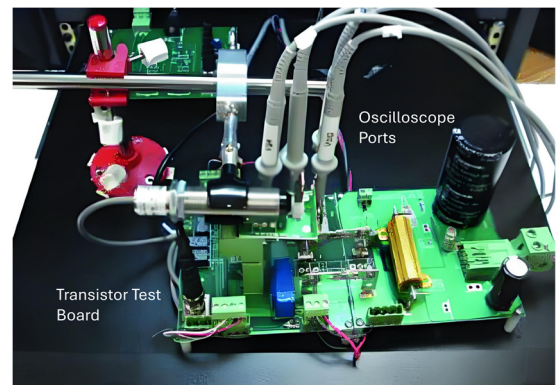


FIGURE 5. IGBT accelerated aging test hardware [33].

MAE is calculated by averaging the absolute differences between predicted and actual values, directly reflecting the average prediction error of the model. Since MAE does not involve squaring the errors, it is less sensitive to outliers, providing a more realistic representation of the overall error level.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

In the equation: n represents the total number of samples; y_i denotes the actual value of the i -th sample; $\hat{y}_i$ denotes the predicted value of the i -th sample.

TABLE 1. Model structure and parameters.

Hyperparameters	Description	Value
lr	Initial learning rate	0.001
scheduler_factor	LR reduction factor	0.5
scheduler_patience	LR patience	5
N_epoch	Train iters	140
Batch_size	Number of batches	64
GRU_hidden_size	GRU hidden units	128
freq_feat_size	Freq feat dim	16
dense_size	FC hidden nodes	32

Table 1 summarizes the hyperparameter settings of the proposed model, including the learning rate, number of training iterations, batch size, and parameter configurations for key network structures.

B. LIFE PREDICTION UNDER DIFFERENT TRAINING STRATEGIES

This experiment further explores the impact of various training strategies on IGBT RUL prediction, specifically comparing enhanced Teacher Forcing strategies (Scheduled Sampling and Professor Forcing) against traditional Teacher Forcing. The experimental results summarized in Table 2 systematically evaluate the influence of each training strategy on prediction accuracy and stability.

The traditional Teacher Forcing strategy accelerates model convergence by consistently feeding the model the true previous outputs as inputs for the subsequent step. However, this method often introduces exposure bias, where prediction errors significantly amplify once the model uses its own previously predicted outputs during inference. The ratio α within Teacher Forcing controls the probability of employing actual data versus model predictions during training. Specifically, when $\alpha = 1.0$, only true values are used, whereas for $\alpha < 1.0$, the model's own predictions are utilized with a probability of $1 - \alpha$. Reducing α encourages the model to gradually learn to correct its own errors but may also increase difficulty in convergence due to early exposure to prediction noise.

Scheduled Sampling addresses this issue by introducing a curriculum learning approach. Initially, training uses a high α value to leverage true data significantly. As training progresses, α is progressively decreased, transitioning the model smoothly towards using more autoregressive predictions, thus preparing it better for the closed-loop environment encountered during inference [22].

Professor Forcing integrates adversarial learning to align the distribution of hidden states between Teacher Forcing and free-running modes. Besides maximizing the likelihood of the next-step prediction, this method introduces a discriminator during training, forcing the network to exhibit indistinguishable internal dynamics between the two operational modes, thus fundamentally eliminating the disparity between training and inference behaviors [23].

Table 2 presents performance metrics (MSE and MAE) across different training strategies, with optimal results highlighted in bold and second-best results underlined. The results indicate that setting the Teacher Forcing ratio to 0.2 yields the best performance, achieving an MSE of 0.0026. This result represents a 38.1% reduction compared to the second-best performance (Teacher Forcing ratio of 0.6, MSE = 0.0042). Additionally, the corresponding MAE at a ratio of 0.2 is 0.045, a 16.7% improvement over the next-best MAE, confirming that lower Teacher Forcing ratios effectively help the model learn long-term dependencies and mitigate cumulative error.

Further analysis reveals that the Teacher Forcing strategy significantly outperforms Curriculum Learning (MSE = 0.0058) and Scheduled Sampling (MSE = 0.0084). Specifically, the MSE improvement is 33.99% compared to Curriculum Learning and 69.05% compared to Scheduled Sampling. Although the MAE improvement against Curriculum Learning (MAE = 0.0558) is modest at 5.50%, it significantly outperforms Scheduled Sampling (MAE = 0.0826), highlighting superior error control capabilities.

It is noteworthy that Curriculum Learning, despite introducing progressive complexity during training, falls short of Teacher Forcing in error control. Scheduled Sampling exhibits notably poorer performance, particularly with elevated MAE values, reflecting its limited effectiveness in accurately predicting drastic changes in degradation curves.

In summary, the Teacher Forcing strategy not only achieves optimal error metrics but also demonstrates considerable advantages in prediction stability and generalization capability. Given its superior performance in capturing key degradation characteristics of IGBTs and enhancing RUL prediction accuracy, this study adopts Teacher Forcing as the training strategy to meet the rigorous precision and stability requirements of practical engineering health-state prediction applications.

TABLE 2. Scores for each model (Optimal results are bolded; suboptimal ones are underlined).

Training strategy	MSE	MAE
Scheduled Sampling	0.0050	0.060
Professor Forcing	0.0043	0.057
Teacher Forcing (ratio=1)	0.047	0.056
Teacher Forcing (ratio=0.8)	0.0054	0.061
Teacher Forcing (ratio=0.6)	0.0042	0.054
Teacher Forcing (ratio=0.4)	0.0073	0.071
Teacher Forcing (ratio=0.2)	0.0026 (38.1%↓)	0.045 (16.7%↓)

C. COMPARATIVE EXPERIMENTS

In this study, 50% of the available historical data was selected as the training dataset, with the trained model subsequently

applied to predict trends in future samples. Besides the proposed GATE model, the study also included a GATE model without the frequency-domain branch, a Bidirectional Gated Recurrent Unit (Bi-GRU), a Transformer model, and a physics-based model as comparison methods to comprehensively evaluate their performances on time-series prediction tasks.

The GRU model effectively handles long-sequence modeling tasks by dynamically selecting and retaining crucial information via its gating mechanism. The Bi-GRU model enhances feature extraction by simultaneously capturing forward and backward sequence features, integrating bidirectional hidden states to provide a more comprehensive understanding of contextual information within the sequence [34]. The Transformer model, characterized by its self-attention mechanism, is capable of modeling dependencies among arbitrary positions within a sequence. It features excellent parallel computation capabilities and scalability, and has been widely applied to tasks such as time-series prediction, classification, and anomaly detection [35].

Additionally, to characterize the degradation trend of the conduction voltage drop of power devices over their usage cycles, this study introduces the following physically inspired exponential growth model [36], [37]:

$$V_{CE}(N) = V_0 + \Delta V_{max} \cdot (1 - e^{-\lambda N^\gamma}) \quad (17)$$

In the equation: $V_{CE}(N)$ is the normalized conduction voltage drop at cycle N ; V_0 represents the reference initial conduction voltage, corresponding to the device's early-state voltage; ΔV_{max} is the maximum voltage drift amplitude during degradation; λ denotes the degradation rate parameter controlling the speed; γ is the nonlinear modulation factor that adjusts the curvature of degradation; and N is the operational cycle number.

This model integrates power-law growth with exponential saturation characteristics, effectively fitting the degradation trajectories of power devices such as IGBTs under thermo-mechanical stress. It is broadly applicable in tasks such as device lifetime evaluation, failure prediction, and health-state modeling.

In the IGBT RUL prediction task, the GATE model demonstrates superior performance across all evaluation metrics compared to other comparative models, indicating exceptional predictive accuracy and robustness. As shown in the table, the MSE of the GATE model is 0.0026, representing a reduction of 38.1% relative to the second-best model, GATE (no freq), which has an MSE of 0.0042. Similarly, the MAE is 0.045, which is 19.6% lower than the 0.056 MAE of GATE (no freq). These results clearly demonstrate that incorporating frequency-domain information while maintaining the original GRU structure significantly enhances overall model performance.

The prediction curves illustrated in Figure 6 further confirm the validity of these metrics. The GATE model (green curve) closely aligns with the true measurement curve (blue)

throughout the entire prediction interval. Particularly during the initial stages of fault evolution, it accurately captures the onset of voltage increases. In contrast, the GRU model (orange) shows delayed responses initially, while Bi-GRU (brown) and Transformer (pink) models exhibit varying degrees of deviation, indicating relatively weaker sensitivity to sequential variations.

During non-stationary phases such as voltage oscillation intervals, the GATE model demonstrates superior stability and noise reduction capability. Its predictions remain smooth and continuous, resisting significant fluctuations from minor disturbances. Conversely, the Transformer model frequently exhibits abrupt jumps and discontinuities, undermining its reliability. This advantage is attributed to the frequency-domain branch integrated into the GATE model, enabling it to effectively identify and suppress high-frequency noise while enhancing sensitivity to characteristic variations.

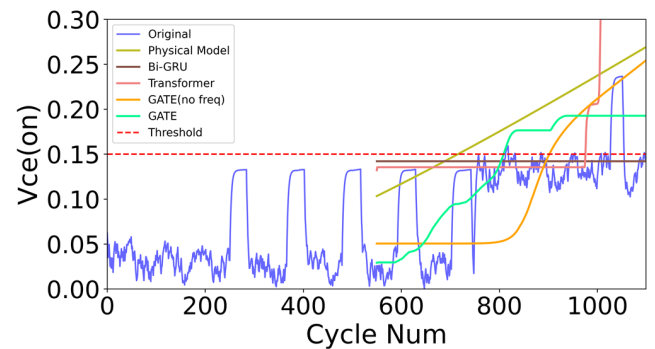


FIGURE 6. Comparison chart of model predictions.

TABLE 3. Scores for each model (Optimal results are bolded; suboptimal ones are underlined).

Model	MSE	MAE
Physical	0.0069	0.073
Transformer	0.0115	0.071
Bi-GRU	0.0047	0.061
GATE(no freq)	0.0042	0.056
GATE	0.0026(38.1%↓)	0.045(19.6%↓)

Furthermore, the GATE model excels near critical points of arc faults, accurately predicting when voltage values approach preset thresholds and proactively responding, thus showcasing robust safety warning capabilities. The Transformer model, on the other hand, exhibits instability around these critical points, with substantial deviations in its predictions, rendering it ineffective as a reliable reference for decision-making.

Overall, compared to conventional GRU, Bi-GRU, and Transformer models, the GATE model not only excels in global fitting accuracy but also significantly improves performance in key aspects such as critical point response, trend capture, noise robustness, and early-warning capabilities. The inclusion of frequency-domain information facilitates

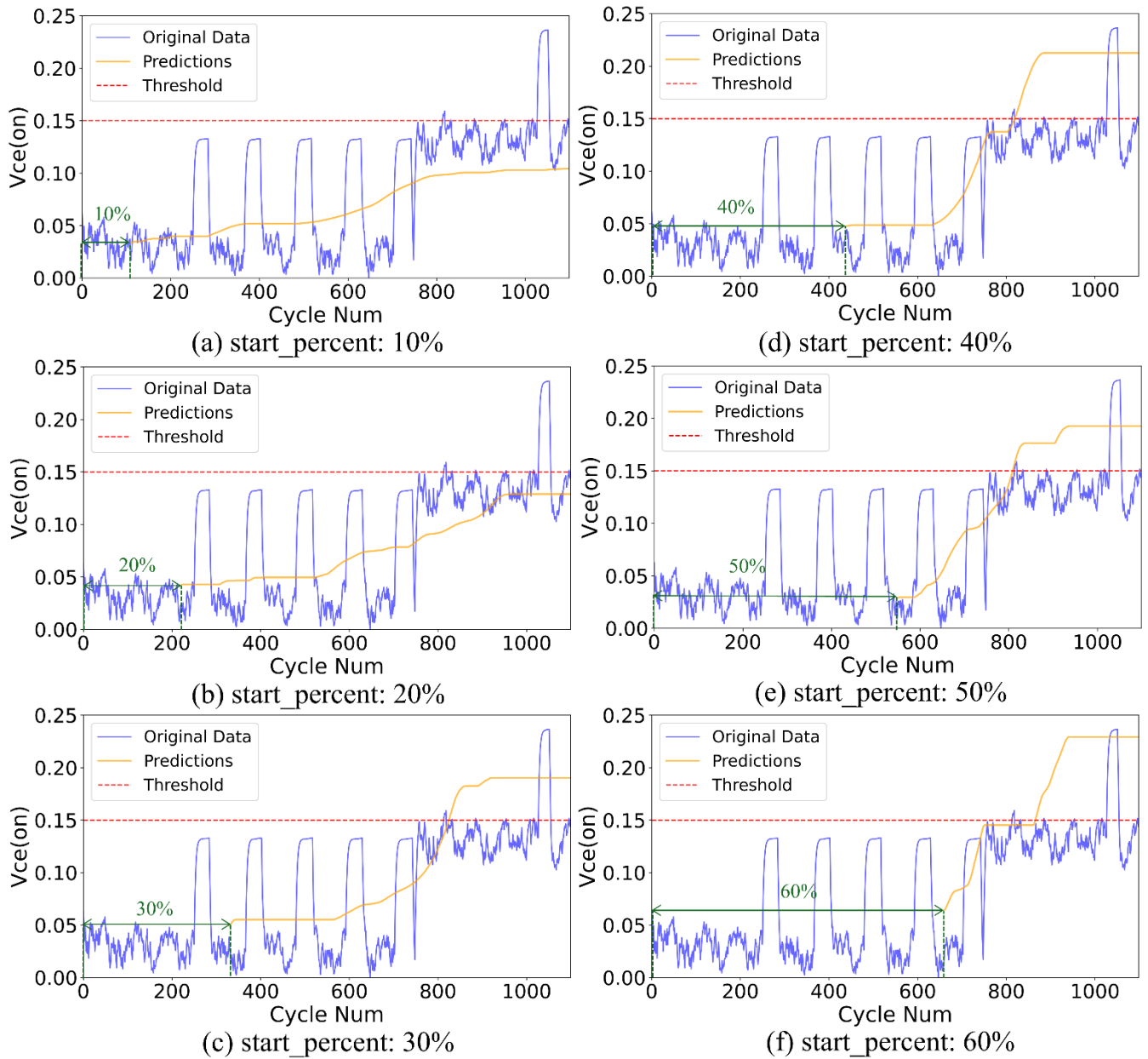


FIGURE 7. Visualization results for different input proportions.

multidimensional feature fusion, particularly suitable for RUL prediction tasks involving electronic devices like IGBTs characterized by complex degradation patterns, demonstrating substantial practical engineering applicability and broader potential for deployment.

D. IMPACT OF DIFFERENT INPUT PROPORTIONS ON PREDICTION

To simulate realistic scenarios where initial historical data availability varies in practical applications, this study evaluated the predictive performance of the GATE model using temperature-compensated data under different proportions of

known historical data. Figure 7 and Table 2 illustrate the impact of varying initial data proportions (start_percent) on prediction outcomes. Experimental results indicate that when the initial data proportion is low (e.g., 10% or 20%), the model can only learn limited degradation characteristics in the early stages, failing to accurately capture significant fluctuations and trends at later stages, thus resulting in notable discrepancies between predictions and actual degradation trajectories.

As the proportion of initial historical data progressively increased (from 10% to 50%), the model's capability to recognize key patterns in the IGBT degradation process

significantly improved, particularly in accurately responding to turning points in degradation trends and high-frequency oscillation nodes. When the input data proportion reached 30% or higher, the GATE model effectively leveraged richer historical information, enhancing its ability to fit degradation trends more accurately. Consequently, the predicted curves closely aligned with the actual data, exhibiting greater stability and sensitivity.

Figure 7 further validates the model's performance across different degradation stages, demonstrating that the GATE model not only effectively models the degradation trajectory but also maintains high prediction accuracy and robustness amidst high-amplitude and high-frequency dynamic changes. These findings highlight its practical value and adaptive flexibility in applications such as power electronics device health assessment and remaining useful life prediction.

E. MODEL PERFORMANCE EVALUATION

To evaluate the runtime efficiency and resource consumption of the proposed GRU model incorporating frequency-domain features in practical deployment scenarios, this study conducted a performance analysis during the inference phase. The experiments were carried out on an NVIDIA GPU-based Linux system platform, utilizing single-sample forward inference. The results indicate that the model achieves an average inference time of 0.8783 milliseconds and a peak GPU memory usage of approximately 15.25 MB, reflecting minimal resource overhead and suitability for online deployment.

The model structure is compact, consisting only of a single-layer GRU and a shallow fully connected network, with moderate computational complexity enhanced by low-dimensional FFT-derived frequency-domain features. Given its uniform input dimensions and independence from complex graph structures, the model also demonstrates strong portability to CPU or low-power embedded GPUs (such as Jetson Nano or NVIDIA Orin Nano). Moreover, the inference time is maintained below 1 ms, meeting real-time requirements typical of power systems and intelligent monitoring systems.

Future work can further compress the model size through quantization (e.g., INT8 inference) or pruning techniques, enabling deployment under more stringent edge computing conditions. Additionally, migration experiments targeting hardware platforms such as FPGAs and DSPs could be conducted to enhance the flexibility and scalability of the model for industrial applications.

V. CONCLUSION AND FUTURE WORK

This research proposes a GATE model integrating time-domain and frequency-domain features to jointly model key degradation characteristics during IGBT aging, aiming to predict the RUL of devices. Unlike traditional neural network approaches reliant solely on time-domain features, the proposed GATE model extracts spectral features from $V_{ce(on)}$ signals using FFT and integrates these with GRU-derived

temporal features. This structure not only enhances accuracy in capturing degradation trends but also provides frequency-domain priors useful for constructing hybrid models that integrate physical mechanisms, aiding in uncovering intrinsic degradation mechanisms under complex operational conditions.

Comparative experiments conducted on publicly available accelerated aging datasets demonstrate that the proposed method achieves superior performance metrics, including MSE and MAE, outperforming traditional temporal models. The notable advantages of incorporating Teacher Forcing mechanisms and frequency-domain features validate the effectiveness of the time-frequency feature integration strategy in enhancing model performance and early degradation sensitivity.

Potential applications of the GATE model include rapid screening prior to device shipment and aging assessment tests at the factory level, enabling accurate life-cycle predictions from early-stage data without the necessity of full lifecycle accelerated testing, thereby significantly reducing testing time and costs. However, this method still has limitations. Primarily, the dataset utilized involves idealized constant duty-cycle power cycling conditions, not fully covering the complex operating modes prevalent in actual industrial systems, such as dynamic load disturbances in motor drives and frequent startups in wind-power inverters. Thus, the current model's robustness and generalization capabilities under complex application scenarios remain insufficiently validated.

Additionally, the GATE model relies on historical $V_{ce(on)}$ data for training and prediction, which in practical deployment may face constraints from data acquisition resolution, measurement noise, and sample scarcity, posing challenges to system stability. Therefore, future research will focus on three main expansion areas: (1) introducing domain-adaptive transfer learning techniques, such as adversarial feature alignment and maximum mean discrepancy (MMD) constraints, to address data distribution shifts between experimental and real industrial domains; (2) employing synthetic degradation data generation methods based on finite element simulations and electrothermal modeling to construct training sample pools under multi-condition and multi-device scenarios, thus improving model adaptability; and (3) collaborating with industrial partners to conduct deployment validation in real application environments, investigating the practicality and maintainability of the GATE model in typical scenarios such as industrial inverters and motor drive systems.

In summary, although GATE demonstrates excellent performance under controlled experimental conditions, systematic resolution of generalization, interpretability, and deployment complexity issues is required for engineering promotion. Future developments integrating multi-source information fusion and cross-domain transfer modeling hold promise for enhancing its application value in intelligent manufacturing and power electronics health management.

ACKNOWLEDGMENT

(Yuankuang Li, Baohua Tan, Yingjie Zhang, and Shuo Wang contributed equally to this work.)

REFERENCES

- [1] A. Abuelnaga, M. Narimani, and A. S. Bahman, "Power electronic converter reliability and prognosis review focusing on power switch module failures," *J. Power Electron.*, vol. 21, no. 6, pp. 865–880, Jun. 2021.
- [2] S. Peyghami, P. Palensky, and F. Blaabjerg, "An overview on the reliability of modern power electronic based power systems," *IEEE Open J. Power Electron.*, vol. 1, pp. 34–50, 2020.
- [3] A. Abuelnaga, M. Narimani, and A. S. Bahman, "A review on IGBT module failure modes and lifetime testing," *IEEE Access*, vol. 9, pp. 9643–9663, 2021.
- [4] T. J. Formica, H. A. Khan, and M. G. Pecht, "The effect of inverter failures on the return on investment of solar photovoltaic systems," *IEEE Access*, vol. 5, pp. 21336–21343, 2017.
- [5] H. Oh, B. Han, P. McCluskey, C. Han, and B. D. Youn, "Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2413–2426, May 2015.
- [6] Y. Yang, H. Wang, A. Sangwongwanich, and F. Blaabjerg, "Design for reliability of power electronic systems," in *Power Electronics Handbook*, M. H. Rashid, Ed., 3rd ed. Oxford, U.K.: Elsevier, 2018, pp. 1423–1440.
- [7] N.-C. Sintamarean, F. Blaabjerg, H. Wang, F. Iannuzzo, and P. de Place Rimmen, "Reliability oriented design tool for the new generation of grid connected PV-inverters," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2635–2644, May 2015.
- [8] A. Watanabe, M. Tsukuda, and I. Omura, "Real-time degradation monitoring system for high power IGBT module under power cycling test," *Microelectron. Reliab.*, vol. 53, nos. 9–11, pp. 1692–1696, 2013.
- [9] X. Fang, S. Lin, X. Huang, F. Lin, Z. Yang, and S. Igarashi, "A review of data-driven prognostic for IGBT remaining useful life," *Chin. J. Electr. Eng.*, vol. 4, no. 3, pp. 73–79, Sep. 2018.
- [10] W. Lai, M. Chen, L. Ran, O. Alatisse, S. Xu, and P. Mawby, "Low stress cycle effect in IGBT power module die-attach lifetime modeling," *IEEE Trans. Power Electron.*, vol. 31, no. 9, pp. 6575–6585, Sep. 2016.
- [11] W. Wu, M. Held, P. Jacob, P. Scacco, A. Birolini, "Investigation on the long term reliability of power IGBT modules," in *Proc. Int. Symp. Power Semiconductor Devices ICS*, pp. 443–448.
- [12] A. Singh, A. Anurag, and S. Anand, "Evaluation of vce at inflection point for monitoring bond wire degradation in discrete packaged IGBTs," *IEEE Trans. Power Electron.*, vol. 32, no. 4, pp. 2481–2484, Apr. 2017.
- [13] M. S. Haque, S. Choi, and J. Baek, "Auxiliary particle filtering-based estimation of remaining useful life of IGBT," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2693–2703, Mar. 2018.
- [14] H. Wu, C. Ye, Y. Zhang, J. Nie, Y. Kuang, and Z. Li, "Remaining useful life prediction of an IGBT module in electric vehicles statistical analysis," *Symmetry*, vol. 12, no. 8, p. 1325, Aug. 2020.
- [15] X. Wang, Z. Zhou, S. He, J. Liu, and W. Cui, "Performance degradation modeling and its prediction algorithm of an IGBT gate oxide layer based on a CNN-LSTM network," *Micromachines*, vol. 14, no. 5, p. 959, Apr. 2023.
- [16] Z. Chen, M. Dai, J. Liu, W. Jiang, and Y. Min, "Life prediction of IGBT module for nuclear power plant rod position indicating and rod control system based on SDAE-LSTM," *Nucl. Eng. Technol.*, vol. 56, no. 9, pp. 3740–3749, Sep. 2024.
- [17] J. Tian, Y. Jiang, H. Luo, and S. Yin, "Adaptive boosting based on multi-class neural networks for IGBT health parameter prediction," in *Proc. 22nd IEEE Int. Conf. Ind. Technol. (ICIT)*, vol. 1, Mar. 2021, pp. 1001–1006.
- [18] S. H. Ali, M. Heydarzadeh, S. Dusmez, X. Li, A. S. Kamath, and B. Akin, "Lifetime estimation of discrete IGBT devices based on Gaussian process," *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 395–403, Jan. 2018.
- [19] C. He, W. Yu, Y. Zheng, and W. Gong, "Machine learning based prognostics for predicting remaining useful life of IGBT-NASA IGBT accelerated ageing case study," in *Proc. IEEE 5th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, vol. 5, Oct. 2021, pp. 1357–1361.
- [20] Y. Wang, F. Xie, T. Zhao, Z. Li, M. Li, and D. Liu, "IGBT status prediction based on PSO-RF with time-frequency domain features," in *Proc. IEEE 11th Data Driven Control Learn. Syst. Conf. (DDCLS)*, Aug. 2022, pp. 337–341.
- [21] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, Jun. 1989.
- [22] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," in *Proc. 29th Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Montreal, QC, Canada, Dec. 2015, pp. 1171–1179.
- [23] A. Lamb, A. Goyal, Y. Zhang, S. Zhang, A. Courville, and Y. Bengio, "Professor forcing: A new algorithm for training recurrent networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 4601–4609.
- [24] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [25] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
- [26] B. Tan, W. You, C. Huang, T. Xiao, S. Tian, L. Luo, and N. Xiong, "An intelligent near-infrared diffuse reflectance spectroscopy scheme for the non-destructive testing of the sugar content in cherry tomato fruit," *Electronics*, vol. 11, no. 21, p. 3504, 2022.
- [27] E. Ahmadzadeh, H. Kim, O. Jeong, N. Kim, and I. Moon, "A deep bidirectional LSTM-GRU network model for automated ciphertext classification," *IEEE Access*, vol. 10, pp. 3228–3237, 2022.
- [28] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
- [29] E. Dimech and J. F. Dawson, "IGBT's ageing and its impacts on the EM conducted emissions," in *Proc. Int. Symp. Electromagn. Compat.-EMC Eur.*, Sep. 2024, pp. 1082–1087.
- [30] Vignesh. (Oct. 2019). *IGBT Accelerated Aging Data Set*. [Online]. Available: <https://www.kaggle.com/datasets/vignesh9147/igbt-accelerated-aging-data-set>
- [31] G. Sonnenfeld, K. Goebel, and J. R. Celaya, "An agile accelerated aging, characterization and scenario simulation system for gate controlled power transistors," in *Proc. IEEE Autotestcon*, Sep. 2008, pp. 208–215.
- [32] S. Yang, D. Xiang, A. Bryant, P. Mawby, L. Ran, and P. Tavner, "Condition monitoring for device reliability in power electronic converters: A review," *IEEE Trans. Power Electron.*, vol. 25, no. 11, pp. 2734–2752, Nov. 2010.
- [33] J. R. Celaya. *Electronics Prognostics*. Accessed: Mar. 2, 2025. [Online]. Available: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/electronics-prognostics/>
- [34] M. K. Hasan, S. A. Islam, M. S. Ejaz, M. M. Alam, N. Mahmud, and T. A. Rafin, "Classifying Bengali newspaper headlines with advanced deep learning models: LSTM, bi-LSTM, and bi-GRU approaches," *Asian J. Res. Comput. Sci.*, vol. 16, no. 4, pp. 372–388, Dec. 2023.
- [35] L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, and B. Huang, "A systematic review for transformer-based long-term series forecasting," *Artif. Intell. Rev.*, vol. 58, no. 3, p. 80, Jan. 2025.
- [36] N. Degrenne and S. Mollov, "Diagnostics and prognostics of wire-bonded power semiconductor modules subject to DC power cycling with physically inspired models and particle filter," in *Proc. PHM Soc. Eur. Conf.*, vol. 4, no. 1, 2018, pp. 1–10.
- [37] J. Wu, Z. Xu, and X. Wei, "Remaining useful life prediction of power MOSFETs using model-based and data-driven methods," in *Proc. Int. Conf. Cyber Secur. Intell. Anal.*, in *Advances in Intelligent Systems Computing*, 2020, pp. 373–381.

YUANKUANG LI is currently pursuing the bachelor's degree in electrical engineering and automation with Detroit Green Industry College, Hubei University of Technology, Wuhan, China. His research interests include electrical systems, machine learning, and photovoltaic systems.





BAOHUA TAN was born in Yingshan, Hubei, China, in 1978. He received the Ph.D. degree in power machinery and engineering from Wuhan University of Technology, Wuhan, China, in 2014.

Since 2016, he has been a Professor with the School of Science, Hubei University of Technology, Wuhan. He is the author of 15 books, more than 160 articles, more than 90 inventions, and holds 80 patents. His research interests include the information exchange technology, computer application technology, measurement, and control technology. He was the Academic Dissertation Evaluation Expert of the Ministry of Education of China, the Evaluation Expert of the Science and Technology Award of Hubei Province, the Project Evaluation Expert of the Environmental Protection Department of Hubei Province, the Evaluation Expert of the Science and Technology Award of Hebei Province, the Evaluation Expert of the Science and Technology Award of Jiangxi Province, and the Expert of Science and Technology Consultation of Wuhan City. He is a Reviewer of several journals, such as *IEEE Access*, *Zhongguo Jixie Gongcheng/China Mechanical Engineering*, *Journal of Sensors*, *Journal of Electronics, Designs*, *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, *Journal of Central China Normal University (Natural Science Edition)*, and *Journal of Hubei University of Technology*.



YINGJIE ZHANG (Member, IEEE) received the bachelor's degree in electrical engineering from Hubei University of Technology. He is currently pursuing the master's degree with the Department of Applied Physics, The Hong Kong Polytechnic University. His research focuses on AI-driven materials discovery and its further applications in the microelectronics industry.



SHUO WANG is currently pursuing the bachelor's degree in software engineering with Detroit Green Industry College, Hubei University of Technology, Wuhan, China. His research interests include machine learning, medical informatization, generative artificial intelligence, and natural language processing.



BOLIN LIAN is currently pursuing the bachelor's degree in software engineering with Detroit Green Industry College, Hubei University of Technology, Wuhan, China. His research interests include machine learning, artificial intelligence, information technology, and big data.



YUAN LI is currently pursuing the bachelor's degree in electrical engineering and automation with Detroit Green Industry College, Hubei University of Technology, Wuhan, China. His research interests include electrical systems, machine learning, and generative artificial intelligence.



ZONGMING TAN was born in Wuhan, Hubei, China, in 2008. His research interests include embedded system development, software design, and robot technology.

...