

Article

Data Augmentation via Auxiliary Classifier GAN for Enhanced Modeling of Gallium Nitride HEMT Devices

Yifei Liu ¹, Yihan Qian ¹, Yefeng Hu ¹ and Ye Wu ^{1,2,*} 

¹ School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China; yifei.liu2202@student.xjtlu.edu.cn (Y.L.); yihan.qian22@student.xjtlu.edu.cn (Y.Q.); yefeng.hu22@student.xjtlu.edu.cn (Y.H.)

² Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK

* Correspondence: ye.wu@xjtlu.edu.cn

Abstract

Accurate and efficient modeling of AlGaIn/GaN HEMTs is essential for the design of next-generation power electronics. This study introduces a hybrid Auxiliary Classifier Generative Adversarial Network (ACGAN)–mixup data augmentation framework to enhance deep neural network application in AlGaIn/GaN high-electron-mobility transistor modeling with limited data. Based on only 20 distinctive devices, ACGAN uses technology computer-aided design (TCAD)-calibrated data to generate high-quality synthetic drain current (I_{ds}) under various electronic bias conditions. The quality of the generated data is validated via Jensen–Shannon divergence with an average of 0.0341. A one-dimensional convolutional neural network (1D-CNN) predictive model is trained on augmented data and achieves stable convergence, with a mean absolute error of 0.002 A/mm for the off-state I_{ds} and 0.052 A/mm for the linear region. It also shows improved robustness over the model trained on original non-augmented data. The proposed approach offers a low-cost alternative to resource-intensive TCAD simulations, enabling accurate device modeling with limited data.

Keywords: Gallium Nitride; semiconductor device; data augmentation; ACGAN

1. Introduction

Gallium Nitride (GaN), as a third-generation semiconductor material, is widely used in power electronic devices due to its superior electrical and material properties compared to conventional silicon (Si)-based devices. These advantages include: a wide bandgap (~ 3.4 eV) enabling higher voltage application [1], high electron mobility (~ 1200 cm²/Vs) achieving faster switching capability [2], and a high breakdown field (~ 3.3 MV/cm) allowing for thinner, more effective devices [3]. However, because of the complexity of the structure and its physics, the behavior of GaN-based devices is still not fully understood, creating barriers in modeling them [4,5].

Neural network (NN)-based models are widely used in scientific areas, such as material science [6], applied physics [7], and biology [8]. But quite a few of their applications are in semiconductor device simulation. Recently, there have been NN applications for GaN-based devices [9–12]. Compared to conventional physics-based modeling methods, these data-driven modeling methods require fewer resources, less professional expertise, and shorter development time. Also, they have high accuracy compared to empirical fitting of the physics-based model parameters. However, foundries and laboratories are still using



Academic Editor: Wojciech Wojtasiak

Received: 22 January 2026

Revised: 19 February 2026

Accepted: 2 March 2026

Published: 4 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

TCAD tools for designing and simulating semiconductor devices. The main reason is that the accuracy of the NN model is highly reliant on the quantity and quality of the data [13], which are hard to determine experimentally. To solve this, TCAD-augmented [14,15] and/or Spice-augmented [16] machine learning (ML) have been used in semiconductor device simulation. Dhillon et al. [15] spent two weeks generating a set of $\sim 150,000$ current-voltage (I-V) curves to train a machine learning model. Tools like Spice and TCAD can generate data in a well-controlled varying environment, but the simulation itself includes approximation and simplification of the actual physical process, which compromises the quality of the data and introduces migration from the real data. It is worth exploring new, effective methods to balance the factors of limited data and high-quality NN models. In recent years, pure data-driven methods such as the Generative Adversarial Network (GAN), variational autoencoders (VAEs), and transforms are widely used as data augmentation methods in the areas of computer vision (CV), natural language processing (NLP), and other scientific areas [11,17–21]. Research applying data-driven augmentation methods for semiconductor simulation is rare. There is still room to explore effective data-driven methods for data augmentation.

In this paper, we propose a hybrid framework that integrates an ACGAN with a mixup method for pure data-driven augmentation to overcome the limited experimental data and high computational expenses of TCAD simulations. We apply an experiment on 20 distinctive GaN HEMT devices and use their electrical performance under different biases as the initial data. The proposed ACGAN model is used to generate physically consistent data with distributions similar to the initial dataset, thereby reducing the need for further TCAD simulations. A 1D-CNN regression model is then employed to predict the drain current (I_{ds}) of GaN HEMT devices based on pure initial data and augmented data, respectively. The augmented data is used to improve the accuracy and efficiency of the 1D-CNN, which has a significant improvement over existing TCAD-augmented and Spice-augmented solutions that require weeks of simulation.

2. Initial Data Generation with TCAD

Silvaco's [22] TCAD simulator was used to generate the initial characteristic data of different GaN HEMT devices. The device was constructed on a silicon carbide substrate with a multilayer epitaxial structure that included a GaN buffer layer, an AlGaIn barrier layer with different compositional parameters, and a GaN cap layer that formed a GaN/Al_xGa_{1-x}N/GaN heterostructure. The surface of the device was covered with a Si₃N₄ passivation layer. The schematic structure of the device is shown in Figure 1. It follows the standard AlGaIn/GaN HEMT architecture reported in ref. [23]. The simulations used models including: mobility models, such as field-dependent carrier mobility and concentration-dependent carrier mobility, and recombination models, including both Shockley–Read–Hall and Auger recombination models to ensure physical accuracy. The Polarization-induced charge effects and deep-level trap states were defined at the Al-GaN/GaN heterointerface to capture charge trapping dynamics. The gate electrode was designed with a work function of 4.3 eV, while both source and drain electrodes were designed as ohmic contacts.

The model was first calibrated using the existing structure in [23] with $L_g = 0.5 \mu\text{m}$, $L_{sg} = 0.5 \mu\text{m}$, and $L_{gd} = 1 \mu\text{m}$. The GaN/Al_xGa_{1-x}N/GaN heterostructure was set to 3 nm/20 nm/600 nm with $x = 0.35$. An additional 50 nm Si₃N₄ passivation layer was on the top of the heterostructure. The calibration was to ensure the physical consistency of our TCAD model prior to generating the varied-device dataset. As shown in Figure 2, the result demonstrated a close match between simulation and experimental outcomes by comparing the transfer characteristics of GaN HEMTs, indicating a valid simulation model.

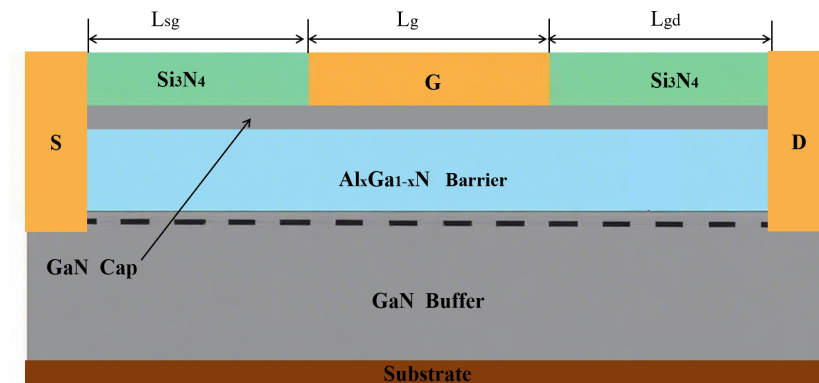


Figure 1. The schematic structure of AlGaIn/GaN HEMT (not drawn to scale).

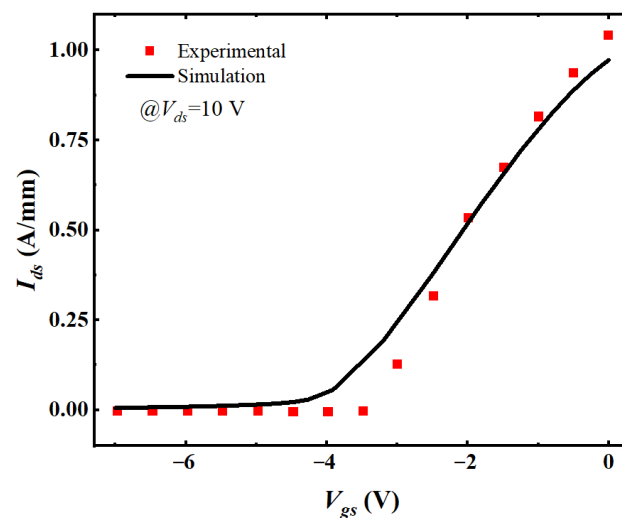


Figure 2. Calibration of TCAD-generated and measured $I_{ds} - V_{gs}$ at $V_{ds} = 10$ V from the experimental AlGaIn/GaN HEMT reported in [23].

The calibrated TCAD model is applied to generate initial data, which is used as a limited dataset in a real application. Different devices are obtained by varying the thickness of the AlGaIn layer (t), the aluminum mole fraction in AlGaIn (%Al), and the doping concentration of the barrier layer (D). For each device, the electrical characteristics are generated by varying the drain voltage (V_{ds}) and the gate voltage (V_{gs}). The ranges and step sizes of each parameter are detailed in Table 1. The selected parameter ranges cover representative design boundaries for AlGaIn/GaN HEMTs, balancing computational feasibility with demonstration of the framework under data-limited conditions similar to those in prior works [9,23]. The parameter sweep provides 24 distinctive devices with unique geometries and material properties and 5544 different drain currents (I_{ds}) under different voltage bias variations. A random selection of 20 devices with 4620 I_{ds} are used for ACGAN-based augmentation, and the remaining four devices are reserved for validation and test purposes.

Table 1. Parameter variation for HEMT device simulation in initial data generation.

Parameters	From	To	Step Size
t (μm)	0.015	0.025	0.005
%Al	15	33	6
D (cm^{-3})	2×10^{18}	6×10^{18}	4×10^{18}
V_{ds} (V)	5	10	0.5
V_{gs} (V)	−5	2	0.35

3. ACGAN-Based Data Augmentation

3.1. Data Preprocessing

The TCAD-generated initial data from GaN HEMT devices included structural parameters (e.g., t), material properties (e.g., %Al, D), bias conditions (e.g., V_{ds} , V_{gs}), and the corresponding drain current I_{ds} . All input features were normalized to a uniform scale of $[0, 1]$ using min–max normalization to ensure numerical stability during both ACGAN and 1D-CNN training. The normalized value eliminated the influence of the huge magnitude difference in different features, aligned all the data on the same scale, and could enhance the overall performance of the models.

3.2. ACGAN Architecture

The proposed ACGAN is designed to generate the full $I_{ds}(V_{gs}, V_{ds})$ response conditioned on physical device parameters, rather than isolated current samples. The label conditions eliminate uncontrolled randomness in the generated outputs and ensure physical consistency with the input parameters. As shown in Figure 3, the ACGAN framework comprises two core components.

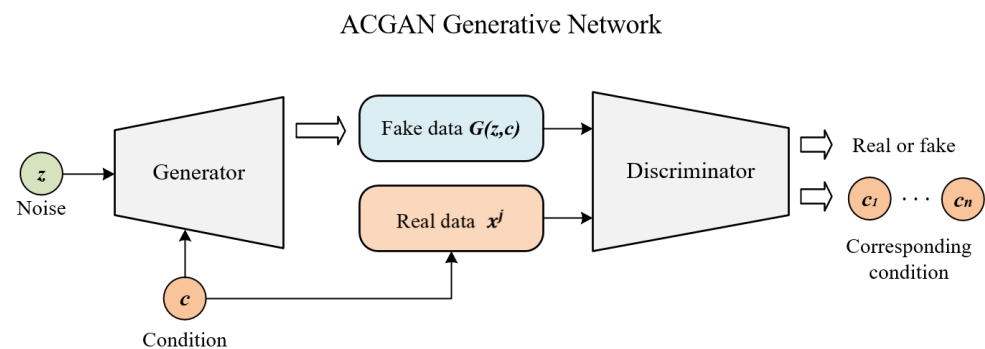


Figure 3. Schematic architecture of the ACGAN framework.

- (1) **Generator (G):** The generator takes a random noise vector z and the conditioning vector c to produce synthetic fake data $G(z, c)$. In this specific implementation, the noise vector z is in Gaussian distribution $\mathcal{N}(0, 1)$ with dimension 128 and the condition vector c is a three-dimensional device design parameter (%Al, t , D). The output fake data $G(z, c)$ is an 11×21 matrix of I_{ds} under discrete V_{ds} and V_{gs} . Since the output has a low dimension, a fully connected multilayer perceptron (MLP) architecture with a ReLU activation functions is used. The fully connected layer has a 1024-512-256 structure as hidden layers.
- (2) **Discriminator (D):** The discriminator is designed to distinguish between real and fake data and to classify the data into the correct category. In this work, the discriminator takes the flattened 11×21 drain current I_{ds} (from TCAD or from the generator) and the condition c as input, which forms a 234-dimensional vector. The fully connected MLP hidden layers for feature extraction have a 1024-5120-256 structure. Each layer applies a non-linear activation function ReLU to capture complex patterns in the data. Batch normalization and dropout (rate of 0.2) are applied after each activation to improve generalization and reduce overfitting. The output is split into two independent heads:
 - Authenticity Network with a sigmoid activation function to give a probability score $D_{src}(I_{ds}, c) \in [0, 1]$ to indicate whether the sample is real or fake. The network aims to evaluate the authenticity of input data I_{ds} and $G(z, c)$.
 - Auxiliary Classifier network with regression head is applied to predict the conditioning information of the data $D_{aux}(I_{ds}, c)$, which is the device information for a given I_{ds} distribution.

Before training, data is fed to the discriminator and the mixup augmentation is applied. For the randomly selected i^{th} and j^{th} samples $(c_i, I_{ds,i})$ and $(c_j, I_{ds,j})$ from the same mini-batch, the new training points are (c_{mix}, I_{mix}) :

$$c_{mix} = \lambda c_i + (1 - \lambda) c_j, \quad I_{ds,mix} = \lambda I_{ds,i} + (1 - \lambda) I_{ds,j} \quad (1)$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$, α is a hyperparameter set to 0.8 and Beta is 0.4 to control the mix ratio. The mixup method creates virtual intermediate devices in the continuous parameter space $(\%Al, t, D)$ and their corresponding electrical responses, addressing the model's imbalance problem [24] and improving the quality of generated samples [25].

3.3. Loss Functions

The objective of the proposed ACGAN is to generate drain current responses $I_{ds}(V_{gs}, V_{ds})$ that are both statistically consistent with TCAD data and physically aligned with the conditioning vector $c = (\%Al, t, D)$. The discriminator is optimized using a combined adversarial loss $\mathcal{L}_{adv}^D$ and auxiliary loss $\mathcal{L}_{aux}^D$:

$$\mathcal{L}_D = \mathcal{L}_{adv}^D + \lambda_{aux} \mathcal{L}_{aux}^D \quad (2)$$

where λ_{aux} is a hyperparameter used to balance the adversarial loss and auxiliary loss. The adversarial term enforces realism of the generated current using the standard Binary Cross-Entropy (BCE) as the loss function:

$$\mathcal{L}_{adv}^D = -\mathbb{E}_{(I_{ds}, c) \sim p_{data}} [\log D_{src}(I_{ds}, c)] - \mathbb{E}_{z \sim \mathcal{N}(0,1), c \sim p_c} [\log(1 - D_{src}(G(z, c), c))] \quad (3)$$

where $D_{src}(I_{ds}, c)$ and $D_{src}(G(z, c), c) \in [0, 1]$ are probability scores indicating whether the sample is real or fake. The auxiliary regression loss enforces consistency between the drain current and their physical device parameters using mean absolute error (L1) as loss function:

$$\mathcal{L}_{aux}^D = \mathbb{E}_{(I_{ds}, c) \sim p_{data}} \|D_{aux}(I_{ds}, c) - c\|_1 + \mathbb{E}_{z, c} \|D_{aux}(G(z, c), c) - c\|_1 \quad (4)$$

where $D_{aux}(I_{ds}, c)$ and $D_{aux}(G(z, c), c)$ are the regression output of the discriminator.

The generator is trained to both produce realistic current and to satisfy the conditional constraint:

$$\mathcal{L}_G = \mathcal{L}_{adv}^G + \lambda_{aux} \mathcal{L}_{aux}^G \quad (5)$$

with:

$$\mathcal{L}_{adv}^G = -\mathbb{E}_{z \sim \mathcal{N}(0,1), c \sim p_c} [\log D_{src}(G(z, c), c)] \quad (6)$$

$$\mathcal{L}_{aux}^G = \mathbb{E}_{z, c} \|D_{aux}(G(z, c), c) - c\|_1 \quad (7)$$

3.4. Training

The ACGAN is trained using an alternating optimization strategy in which the generator and discriminator are updated iteratively to learn the joint distribution between device conditions and drain current response. Training is performed for 2000 epochs with a batch size of 64, and mixup with a coefficient of 0.2 is applied to interpolate between samples and improve coverage of the device design space. A simplified pseudocode of the training procedure is shown in Algorithm 1.

Algorithm 1: Training procedure for ACGAN–mixup framework.

Input : TCAD-calibrated dataset $\mathcal{X}_{real}$; Device parameters $\mathbf{C}_{real}$;
Hyperparameters: balancing weight λ_{aux} , mixup coefficient α , batch size m , learning rate η , number of epochs N .

Output: Optimized Generator G for GaN HEMT I_{ds} synthesis.

- 1 Initialize weights θ_G and θ_D for Generator G and Discriminator D ;
- 2 **for** $epoch = 1$ **to** N **do**
- 3 **for** each mini-batch (x_i, c_i) in $\mathcal{X}_{real}$ **do**
- 4 // Data Augmentation via Mixup
Sample $\lambda \sim \text{Beta}(\alpha, \alpha)$ and generate x_{mix}, c_{mix} by interpolating (x_i, c_i) and (x_{i+1}, c_{i+1}) ;
- 5 // Update Discriminator D
Generate fake samples $x_{fake} = G(z, c_{mix})$;
- 6 $\mathcal{L}_{adv} = \text{BCE}(D_{src}(x_{mix}), 1) + \text{BCE}(D_{src}(x_{fake}), 0)$;
- 7 $\mathcal{L}_{aux} = \text{MAE}(D_{aux}(x_{mix}), c_{mix}) + \text{MAE}(D_{aux}(x_{fake}), c_{mix})$;
- 8 $\mathcal{L}_D = \mathcal{L}_{adv} + \lambda_{aux} \mathcal{L}_{aux}$;
- 9 $\theta_D \leftarrow \text{Adam}(\nabla_{\theta_D} \mathcal{L}_D, \theta_D, \eta)$;
- 10 // Update Generator G
Sample new noise $z \sim \mathcal{N}(0, 1)$ and generate $x_{fake} = G(z, c_{mix})$;
- 11 $\mathcal{L}_{adv} = \text{BCE}(D_{src}(x_{fake}), 1)$;
- 12 $\mathcal{L}_{aux} = \text{MAE}(D_{aux}(x_{fake}), c_{mix})$;
- 13 $\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{aux} \mathcal{L}_{aux}$;
- 14 $\theta_G \leftarrow \text{Adam}(\nabla_{\theta_G} \mathcal{L}_G, \theta_G, \eta)$;
- 15 **if** Converged **then**
- 16 Break training;

3.5. Post-Generation Data Validation and Cleaning

To further determine whether the data generated by the generative model is realistic, this work uses Jensen–Shannon (JS) divergence to quantify the similarity between the synthesis data and the original data. JS divergence is defined as

$$JS(P||Q) = \frac{1}{2} D_{KL}(P||\frac{(P+Q)}{2}) + \frac{1}{2} D_{KL}(Q||\frac{(P+Q)}{2}), \quad (8)$$

where P and Q are the two probability distributions of the two compared data distributions. D_{KL} is the Kullback–Leibler divergence. If the value of $JS(P||Q)$ is close to 0, that indicates a higher degree of similarity between the distributions [26,27]. To evaluate the quality of the generated data, the probability distribution Q is defined as the distribution of the original data from the TCAD of the 20 HEMT devices. P is defined as the set of newly generated data.

K-means clusters are used for data cleaning. The conditioning vector c (%Al, t , and D) and the flattened I_{ds} vector are embedded as a joint space. K-means is performed on this space, and the centroid and Euclidean distance are computed for each cluster to remove the nonphysical characteristics or poor conditional alignment.

4. Prediction Model Design

A prediction model is designed and trained to estimate the electrical performance of the GaN HEMT devices in order to substitute for the time-consuming device modeling

and simulation process. The input is the structural parameters, material properties, and bias, which form a five-dimensional vector ($\%Al$, t , D , V_{gs} , and V_{ds}). The output is drain current I_{ds} . The training data is normalized before being fed to the prediction model. To better extract the non-linear relation between the input parameters, the 1D-CNN is used. The optimized network consists of an input layer accepting the five-dimensional input data, followed by two fully connected layers (64 and 128 units with ReLU activation), three 1D-convolutional blocks (filter sizes of 32-16-16, kernel size of 3, ReLU activation, batch normalization, and 20% dropout), a global average pooling layer, and a final dense layer with linear output to produce the normalized I_{ds} . Mean Squared Error (MSE) is used as a loss function, which is defined as follows:

$$MSE(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (9)$$

where y_i is the training current value, $\hat{y}_i$ is the predicted value, and N is the number of values. The 1D-CNN is trained using mini-batch gradient-based optimization. During each iteration, batches are randomly sampled from the available dataset and network parameters are updated by minimizing the MSE loss. Early stopping based on validation performance is applied to prevent overfitting.

To evaluate the impact of the proposed data augmentation framework on surrogate modeling performance and ensure a fair and consistent comparison, two independent 1D-CNN predictors are trained using identical network architectures and initialization and optimization settings. One model is trained on the 20 original TCAD samples, while the other is trained on the combined dataset consisting of the original data and the augmented data. The same set of held-out original TCAD samples is used to validate the two models. Performance is quantified by computing the MSE for each training run.

5. Results and Discussion

This section presents the assessment of the ACGAN-based data augmentation framework for GaN HEMT modeling. The analysis includes three aspects: first, the stability of the adversarial training process and the quality of generated samples; second, the impact of the augmented dataset on the predictive performance of GaN HEMT modeling; and finally, broader implications are discussed in relation to conventional approaches. All quantitative evaluations in this section are performed using TCAD-simulated GaN HEMT data as the reference ground truth.

5.1. ACGAN Training Stability and Data Quality

ACGAN is employed to augment the initial dataset and to systematically assess the impact of data augmentation on the performance of prediction models. The ACGAN model is trained and fine-tuned on the hyperparameter; the learning curve is shown in Figure 4, where the loss of the generator and discriminator learning curves are gradually convergent to around 0.46 and 0.48, respectively, after over 800 epochs. Specifically, the standard deviations of the generator and discriminator losses from epoch 800 to 1500 are 0.008 and 0.007, respectively, corresponding to loss fluctuations of less than 2% for both.

Before using the generated data in device modeling, the distribution of the synthesis data is evaluated and K-means clustering is applied to remove sparsely distributed data and outliers. For each cluster, the centroid is computed, and samples whose Euclidean distance from their assigned centroid exceeds a threshold of 1.8 times the cluster's mean radius are discarded. The physically impossible data, such as a negative device value and an abrupt change in current, are also filtered to ensure the physical consistency of the synthesis data with the initial TCAD-generated data.

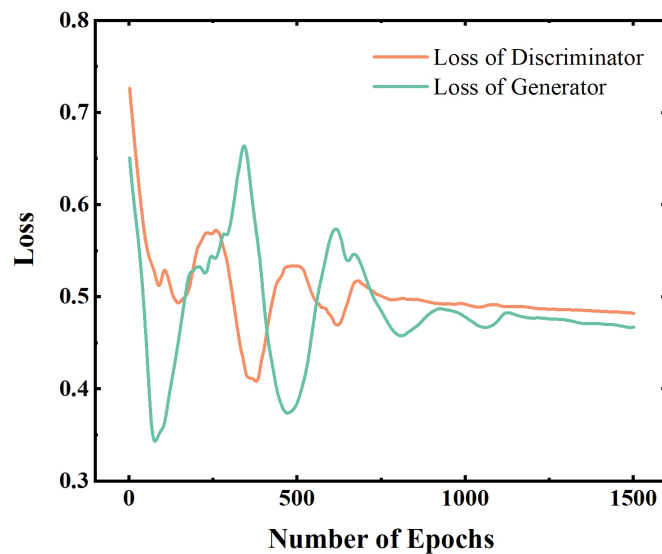


Figure 4. Learning curve of ACGAN.

The filtered data is further confirmed by a low JS divergence of 0.034 ± 0.003 from the initial distribution. As shown in Figure 5, the averaged JS distance is calculated with respect to the initial 20 devices. For each sample, low JS values indicate a high degree of similarity in distribution between the generated data and the initial data. This also indicates the success in capturing the latent characteristics of the data.

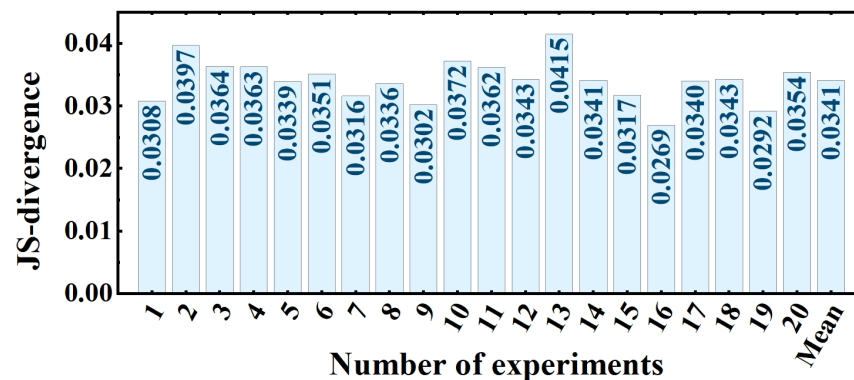


Figure 5. Averaged JS divergence of the generated data for 20 devices.

5.2. Impact of Data Augmentation on Predictive Performance

After filtering, 83,352 drain current I_{ds} samples with different %Al, t , D , V_{gs} , and V_{ds} generated by the ACGAN, together with the initial data, are used to train the 1D-CNN model proposed in Section 4. The untouched four distinctive devices with 924 I_{ds} are used as a validation set. The learning curve of training and validation on the augmented dataset is shown in Figure 6. After more than 1000 epochs, the loss of training and validation becomes stable and convergent. The MSE of approximately 0.226 on validation data demonstrates high accuracy and robustness in predicting I_{ds} for GaN HEMT devices under varying bias conditions. In contrast, training the same 1D-CNN architecture solely on the original dataset from 20 distinctive devices (comprising 4620 I_{ds} under different biases) leads to unstable behavior. The training outcome varies with different initializations and learning rates, and the model fails to converge reliably, exhibiting erratic loss curves.

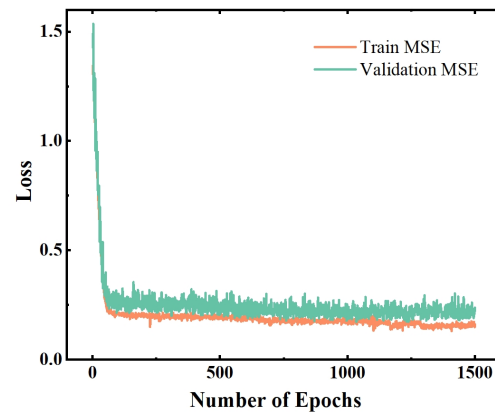


Figure 6. MSE of the optimized 1D-CNN model at each epoch of ACGAN-generated data.

Figure 7 shows representative comparisons of ground-truth TCAD-simulated drain current I_{ds} (solid line) and 1D-CNN-predicted I_{ds} (dots) for unseen validation devices. According to the result, in the off-state region ($V_{gs} < \text{threshold}$, where I_{ds} is very low, usually less than 0.1 A/mm), the mean absolute error remains as small as 0.002 A/mm, preserving the device's low-leakage behavior with high fidelity for accurate estimation. In the linear region (near-threshold to moderate V_{gs} , where I_{ds} increases linearly with V_{gs}) the mean absolute error is 0.052, with the largest absolute error being less than 0.1 A/mm. These regime-specific results demonstrate that the ACGAN-augmented training enables physically meaningful I_{ds} predictions, supporting the framework's utility as an efficient surrogate for early-stage GaN HEMT design.

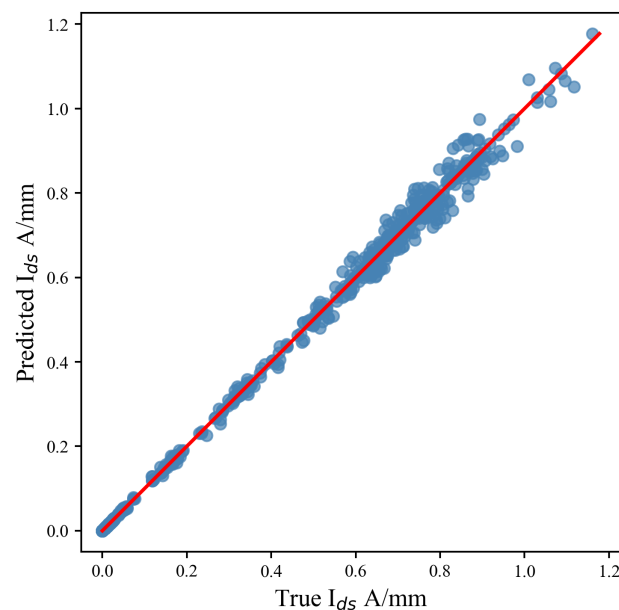


Figure 7. 1D-CNN prediction of I_{ds} : Comparison between ground truth (solid line) and predicted values (dots).

As shown in the experiment, without sufficient data, deep learning-based device prediction becomes unreliable, suffering from non-convergence, severe overfitting, and poor generalization to unseen configurations. The proposed ACGAN framework effectively addresses this limitation by providing a sufficiently rich and diverse training dataset. With the augmented data, a stable surrogate model with low MSE can be obtained, enabling accurate and efficient device prediction. This work requires only a few hours for training and the trained model can complete device prediction in seconds, which is significantly

more efficient than the traditional TCAD or SPICE modeling that may require weeks of simulation [15]. The proposed method is purely data-driven, requires fewer specialized physics-domain skills than TCAD tuning, and is designed with a flexible conditioned architecture that could, in principle, be adapted to other FET-like devices in future work (pending validation on additional device families).

To further evaluate the effectiveness of the proposed data augmentation framework, a controlled comparative study is conducted using representative baseline methods, such as standard GAN and ACGAN without mixup. For all methods, the amount of generated data is kept at the same order of magnitude, and identical data-filling methods are used to ensure a fair comparison. Each augmented dataset is used to train the same 1D-CNN surrogate model, and the resulting prediction performance is evaluated on the same TCAD validation set. The quantitative comparison of these data augmentation strategies for GaN HEMT surrogate modeling is summarized in Table 2.

Table 2. Quantitative comparison of different data augmentation strategies for GaN HEMT surrogate modeling using the same 1D-CNN architecture and TCAD validation set.

Data Augmentation Method	Mean Absolute Error of I_{ds} (A/mm)	
	Off-State	Linear Region
No augmentation (original TCAD only)	Model not converged	
Standard GAN	0.021	0.095
ACGAN (without mixup)	0.015	0.088
Proposed ACGAN–mixup	0.002	0.052

The comparative results in Table 2 demonstrate that all augmentation methods improve performance over the no-augmentation baseline (where the model fails to converge), with the proposed ACGAN–mixup achieving the lowest mean absolute errors of 0.002 A/mm in the off-state and 0.052 A/mm in the linear region. Standard GAN and ACGAN without mixup also yield relatively small absolute errors (0.021/0.015 A/mm in off-state and 0.095/0.088 A/mm in linear region), but these values are misleading in the off-state: since true I_{off} is extremely low in conventional GaN HEMTs, even modest absolute errors translate to high relative/percentage errors, limiting practical utility for leakage assessment. In contrast, the linear region shows more acceptable results across methods, with errors capturing essential current trends reasonably well. The poorer performance of standard GAN originates from its lack of conditioning, resulting in uncontrolled generation that loses device-specific physical controllability. ACGAN without mixup benefits from auxiliary conditioning for better parameter alignment, yet still suffers from the discrete, sparse nature of the 20-device dataset, leading to insufficient smoothing of the data manifold and reduced generalization; mixup addresses this by creating continuous interpolations, ultimately yielding the most robust and accurate predictions.

5.3. Implications and Advantages

In most real applications of semiconductor device modeling, especially for GaN HEMT devices, data limitation is a common problem. This is due to high fabrication costs, long measurement cycles, and limited access to foundry processes. The reliable samples that can be used for training are far below being enough to get a robust predictive model. Such data scarcity leads to critical issues, including training instability, non-convergence, and severe overfitting, as observed in the real-data-only training process in this study. Without sufficient diversity in the training set, the surrogate model is unable to learn generalizable mappings between device parameters and electrical characteristics, not even for new configurations or design iterations. The proposed ACGAN-based augmentation

framework effectively overcomes these limitations by generating a large volume of high-fidelity synthetic I_{ds} - V_{gs} curves or I_{ds} - V_{ds} (83,352 distinct I_{ds} after post-generation cleaning) for different physical parameters as the original data. The augmented dataset enables stable convergence of the 1D-CNN regressor and achieves a substantial reduction in prediction error. For the unseen test sets, the trained model also shows robust generalization. While promising within the explored parameter subspace, extrapolation to denser or broader design spaces requires further validation in future work. Traditional TCAD simulation takes weeks for full-physics simulations, and still introduces approximations that deviate from actual device behavior. In contrast, the ACGAN framework requires only a small set of calibrated samples for training and generates thousands of physically consistent synthetic curves in a few hours. Specifically, ACGAN training requires ~ 3.8 h on an NVIDIA RTX 4080 GPU (16 VRAM), after which synthetic I_{ds} generation is near-instantaneous, in contrast to the days to weeks required for equivalent TCAD parameter exploration. Moreover, the trained model is device-parameter-based and easy to use and set up, thereby lowering the barrier to entry for researchers and engineers in both academic and industrial settings. This approach has the potential to accelerate design cycles, reduce development costs, and facilitate rapid prototyping in GaN-based power electronics. It replaces resource-intensive simulation workflows in next-generation semiconductor device research and development.

6. Conclusions

This study presents a robust hybrid framework integrating the ACGAN with mixup data augmentation, addressing the critical challenge of data scarcity in AlGaIn/GaN HEMT modeling. With only 20 TCAD-calibrated distinct devices, the proposed ACGAN architecture generates synthetic drain current I_{ds} . The screened synthetic data is validated by a low average JS divergence of 0.0341, which indicates consistency with real-world measurements. When evaluating the performance of a 1D-CNN predictive model, with the ACGAN framework augmented data, the mean absolute error of the off-state I_{ds} is 0.002 A/mm and that of linear region is 0.052 A/mm, respectively. The robustness of the model is also improved compared to counterparts trained exclusively on raw, limited data. This approach has the potential to replace resource-intensive, time-consuming conventional modeling methods and provide a cost-effective approach. The findings of ACGAN-based data augmentation on GaN HEMT device modeling solve the data limitation constraints in data-driven model application in the area of semiconductors. It has high predictive accuracy, enabling accelerated design cycles and optimized device performance. This work provides a valuable methodological framework for future research where experimental data acquisition is impractical or costly and paves the way for more scalable, data-efficient surrogate modeling in GaN HEMT applications and, potentially, other wide-bandgap transistor families, subject to further validation across diverse device structures.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, Y.L., Y.Q. and Y.H.; validation, Y.L., Y.Q. and Y.H.; formal analysis, Y.L.; investigation, Y.Q.; resources, Y.W.; data curation, Y.Q.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L., Y.Q. and Y.H.; visualization, Y.L., Y.Q. and Y.H.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the XJTLU Research Development Funding RDF-21-02-072.

Data Availability Statement: Data is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Sharma, R.C.; Nandal, R.; Tanwar, N.; Yadav, R.; Bhardwaj, J.; Verma, A. Gallium arsenide and gallium nitride semiconductors for power and optoelectronics devices applications. *Proc. J. Phys. Conf. Ser.* **2023**, *2426*, 012008. [CrossRef]
- Udabe, A.; Baraia-Etxaburu, I.; Diez, D.G. Gallium nitride power devices: A state of the art review. *IEEE Access* **2023**, *11*, 48628–48650. [CrossRef]
- Zhang, H.; Dong, Y.; Bi, C.; Ding, K.; Zhou, T.; Zhou, H.; Zhou, J.; Zhang, Y. High-Power-Handling GaN Terahertz High-Speed On-Chip Modulator Based on Fin-Line Double-Layer Near-Field Coupling. *IEEE Electron Device Lett.* **2025**, *46*, 1521–1524. [CrossRef]
- Jones, E.A.; Wang, F.F.; Costinett, D. Review of commercial GaN power devices and GaN-based converter design challenges. *IEEE J. Emerg. Sel. Top. Power Electron.* **2016**, *4*, 707–719. [CrossRef]
- Huang, X.; Liu, Z.; Li, Q.; Lee, F.C. Evaluation and application of 600 V GaN HEMT in cascode structure. *IEEE Trans. Power Electron.* **2013**, *29*, 2453–2461. [CrossRef]
- Schmidt, J.; Marques, M.R.; Botti, S.; Marques, M.A. Recent advances and applications of machine learning in solid-state materials science. *NPJ Comput. Mater.* **2019**, *5*, 83. [CrossRef]
- Baldi, P.; Bauer, K.; Eng, C.; Sadowski, P.; Whiteson, D. Jet substructure classification in high-energy physics with deep neural networks. *Phys. Rev.* **2016**, *93*, 094034. [CrossRef]
- Wang, X.; Alnabati, E.; Aderinwale, T.W.; Subramaniya, S.R.M.V.; Terashi, G.; Kihara, D. Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning. *Biophys. J.* **2021**, *120*, 81a. [CrossRef]
- Mishra, S.; Chaturvedi, N. Multi-output deep learning model for simultaneous prediction of figure of merits (Ion, Gm, and Vth) of gallium nitride high electron mobility transistors. *J. Appl. Phys.* **2022**, *131*, 064901. [CrossRef]
- Jiang, Z.; Jiang, Y.; Chen, M.; Li, J.; Li, P.; Chen, B.; Zhao, S.; Wang, J.; Jiang, S.; Cai, M.; et al. Advanced design of a III-nitride light-emitting diode via machine learning. *Laser Photonics Rev.* **2023**, *17*, 2300113. [CrossRef]
- Wang, Z.; Li, L.; Leon, R.C.; Yang, J.; Shi, J.; van der Laan, T.; Usman, M. Improving semiconductor device modeling for electronic design automation by machine learning techniques. *IEEE Trans. Electron Devices* **2024**, *71*, 263–271. [CrossRef]
- Luo, H.; Yan, X.; Zhang, J.; Guo, Y. A neural network-based hybrid physical model for GaN HEMTs. *IEEE Trans. Microw. Theory Tech.* **2022**, *70*, 4816–4826. [CrossRef]
- Hirtz, T.; Huurman, S.; Tian, H.; Yang, Y.; Ren, T.L. Framework for TCAD augmented machine learning on multi-I-V characteristics using convolutional neural network and multiprocessing. *J. Semicond.* **2021**, *42*, 124101. [CrossRef]
- Mehta, K.; Wong, H.Y. Prediction of FinFET current-voltage and capacitance-voltage curves using machine learning with autoencoder. *IEEE Electron Device Lett.* **2020**, *42*, 136–139. [CrossRef]
- Dhillon, H.; Mehta, K.; Xiao, M.; Wang, B.; Zhang, Y.; Wong, H.Y. TCAD-augmented machine learning with and without domain expertise. *IEEE Trans. Electron Devices* **2021**, *68*, 5498–5503. [CrossRef]
- Lu, T.; Kanchi, V.; Mehta, K.; Oza, S.; Ho, T.; Wong, H.Y. Rapid mosfet contact resistance extraction from circuit using spice-augmented machine learning without feature extraction. *IEEE Trans. Electron Devices* **2021**, *68*, 6026–6032. [CrossRef]
- Chen, J.; Wu, Y.; Jia, C.; Zheng, H.; Huang, G. Customizable text generation via conditional text generative adversarial network. *Neurocomputing* **2020**, *416*, 125–135. [CrossRef]
- Gao, X.; Xiao, Z.; Deng, Z. High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *J. Food Eng.* **2024**, *365*, 111833. [CrossRef]
- Duan, X.; Liu, J.; Zhang, E. Efficient image encryption and compression based on a VAE generative model. *J. -Real-Time Image Process.* **2019**, *16*, 765–773. [CrossRef]
- Hossain, M.Z.; Goyal, S. Advancements in Natural Language Processing: Leveraging Transformer Models for Multilingual Text Generation. *Pac. J. Adv. Eng. Innov.* **2024**, *1*, 4–12. [CrossRef]
- Park, T.; Kwak, J.; Ahn, H.; Lee, J.; Lim, J.; Yu, S.; Shin, C.; Moon, T. GAN-based framework for unified estimation of process-induced random variation in FinFET. *IEEE Access* **2022**, *10*, 130001–130023. [CrossRef]
- Guichard, E.; Silvaco, I. Silvaco TCAD, 2022. Available online: <https://nanohub.org/resources/silvacotcad> (accessed on 11 October 2024).
- Raja, P.V.; Nallatamby, J.C.; DasGupta, N.; DasGupta, A. Trapping effects on AlGaIn/GaN HEMT characteristics. *Solid-State Electron.* **2021**, *176*, 107929. [CrossRef]
- Jin, X.; Zhu, H.; Li, S.; Wang, Z.; Liu, Z.; Tian, J.; Yu, C.; Qin, H.; Li, S.Z. A Survey on Mixup Augmentations and Beyond. *arXiv* **2025**, arXiv:2409.05202. [CrossRef]
- Ma, Y.; Xu, X.; Li, Y. Lung RN+ NL: An Improved Adventitious Lung Sound Classification Using Non-local Block ResNet Neural Network with Mixup Data Augmentation. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020; pp. 2902–2906.

26. Shen, X.; Chen, K.; Zhang, T. Asymptotic Statistical Analysis of f -divergence GAN. *arXiv* **2022**, arXiv:2209.06853. [[CrossRef](#)]
27. Shannon, M.; Poole, B.; Mariooryad, S.; Bagby, T.; Battenberg, E.; Kao, D.; Stanton, D.; Skerry-Ryan, R. Non-saturating GAN Training as Divergence Minimization. *arXiv* **2020**, arXiv:2010.08029. [[CrossRef](#)]

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