

Modeling of Magnetic Core Losses From Global to Local Scales Based on Multimodal Neural Networks

Chuangyang WANG, Shutong SANG, Sheng QU, Dongxing YANG, and Yuqi WEI

Abstract—This paper presents a comprehensive framework for modeling magnetic core losses, validated through a systematic multi-scale approach. The framework is built on two pillars: a suite of physically-grounded data augmentation and training optimization techniques, and a multimodal neural network that synergizes CNNs and Transformers. Data augmentation techniques—including random period shifts, waveform flipping, noise injection, and parameter perturbations—simulate real-world waveform irregularities, enabling direct applicability to arbitrary-phase and non-periodic inputs. Complementary training strategies ensure stable learning. The core network architecture dynamically fuses scalar parameters and sequential waveforms via attention mechanisms. Extensive validation uses both an in-house experimental platform and the public MagNet database, extending beyond common ferrites to include nanocrystalline and iron powder cores, with data for some materials reaching magnetic saturation. The framework generalizes consistently across these materials, with average prediction errors below 3.4% for most, and handles non-integer periodic waveforms like 3C94 with about 2.44% error. Compared to existing methods, the core contribution lies in the holistic framework integrating expanded experimental data, enhancement techniques tailored to real-world variability, and rigorous multi-material validation, offering a flexible and directly applicable solution for predicting core losses in power electronic devices.

Index Terms—Core loss, global validation, multimodal fusion networks, self-attention transformer.

I. INTRODUCTION

WITH the rapid advancement of power electronics technology, magnetic components such as inductors and transformers play a critical role in various power electronic systems. Due to the characteristics of magnetic components, such as high-power dissipation and relatively large volume, their design is crucial for improving system efficiency, reducing energy consumption, and achieving structural lightweighting [1]–[3]. Core loss is one of the primary causes of loss in magnetic devices. However, compared to breakthroughs in circuit topology and power devices, current research in magnetic device modeling and design remains relatively underdeveloped. Core

losses are influenced by various nonlinear factors, such as magnetic flux density, excitation frequency, operating temperature, and core material type, presenting significant challenges for loss modeling [4]. Therefore, research on efficient and accurate loss modeling for magnetic cores holds significant importance for the entire field of power electronics.

Currently, the Steinmetz equation and its modified forms are commonly used to model and predict magnetic core losses [5]–[7]. While these quasi-analytical empirical methods provide reasonable estimates within certain ranges, they exhibit significant limitations when confronted with complex scenarios such as non-sinusoidal waveforms, high-frequency excitation, and cross-coupling effects.

In recent years, significant progress has been made in data-driven modeling research. Neural networks demonstrate distinct advantages in handling complex nonlinear mapping problems and exhibit strong universality, leading to their extensive application in magnetic core loss modeling [8], [9]. Current conventional neural network modeling methods typically accept scalars or combinations of scalars and sequences to predict core losses. Simple model architectures struggle to capture the intrinsic relationship between complete inputs and outputs [10], [11]. Consequently, complex model architectures have recently been developed to enable precise predictions.

In [12], researchers proposed a deep neural network (DNN) approach for estimating magnetic core losses. By comparing various internal parameters of the DNN neural network through testing, they determined the optimal DNN structure for magnetic core loss estimation. Researchers proposed a knowledge-aware artificial neural network (KANN) in [13] and applied it to high-frequency magnetic core loss modeling using an improved generalized Steinmetz equation as supplementary knowledge, thereby enhancing both accuracy and speed. These methods are generally only capable of predicting magnetic induction intensity waveforms under relatively regular conditions. Furthermore, they are primarily designed for predicting specific materials, and their models have not been verified for universality across multiple materials, thus exhibiting certain limitations.

In data-driven modeling, data preprocessing and input formats are critical to data quality and prediction accuracy. Consequently, data augmentation strategies can help models exhibit superior generalization capabilities and robustness during prediction [14]. However, this approach also introduces corresponding challenges, such as difficulty in balancing weight distributions and high computational resource demands. Fur-

Manuscript received January 28, 2026; revised March 25, 2026 and April 21, 2026; accepted May 17, 2026. Date of publication September 30, 2026; date of current version July 15, 2026. No funding was received to assist with the preparation of this manuscript. (Corresponding author: Yuqi Wei.)

All authors are with Xi'an Jiaotong University, Xi'an 710049, China (email: 2221310048@stu.xjtu.edu.cn; shutongsang@stu.xjtu.edu.cn; 654321@stu.xjtu.edu.cn; 2212212992@stu.xjtu.edu.cn; yuqiwei@xjtu.edu.cn).

Digital Object Identifier 10.24295/CPSSPEA.2026.00025

thermore, considering that magnetic induction intensity waveforms contain numerous latent features—such as the slope of sine waves, the spike characteristics of triangular waves, and the three-segment structure of square waves—all of which are potential factors influencing the final core loss, feeding the complete waveform sequence into neural network models for prediction can comprehensively capture these latent features. This approach also eliminates the need for additional feature extraction, enhancing convenience while enriching multimodal features and further improving the performance of core loss modeling.

Furthermore, extending the modeling framework to encompass various magnetic materials such as ferrites, nanocrystalline materials, and magnetic powder cores is crucial. Different core materials exhibit markedly distinct electromagnetic properties and loss mechanisms under varying frequencies, temperatures, and excitation waveforms [15]. For instance, ferrites are renowned for their low-loss characteristics at high frequencies, nanocrystalline materials offer superior permeability and thermal stability, while powder cores are valued for their high saturation flux density [16]–[18]. A data-driven model capable of adapting to diverse materials can significantly enhance its practicality in power electronics design across multi-material scenarios.

Based on the above research and discussion, this paper proposes a modeling method for magnetic core losses, expands the dataset to cover both a wide temperature range and a wide frequency range, and employs a multimodal neural network to validate the model using a top-down approach from the global to the local level. The contributions of the proposed modeling method can be divided into two main parts.

First, this paper proposes a CNN-transformer-attention mechanism neural network model. This approach combines the advantages of convolutional neural networks in extracting local waveform features with the global sequence feature observation capabilities of transformers. It integrates waveform and parameter information through a multi-head attention mechanism to achieve accurate prediction of core losses. To enhance the model's generalization capability and training efficiency, data augmentation and training optimization strategies are adopted. These mitigate issues such as overfitting and gradient explosion inherent in traditional data-driven models due to reliance on training data and excessive iteration. The model also adapts to magnetic flux density sequences with arbitrary phases and non-integer period lengths. Using 3C94 material as an example, the model maintains prediction errors below 3.3% for non-integer period flux density sequences, achieving an average percentage error of 2.44%.

Second, the other key contribution of this paper lies in expanding the engineering applicability and functional boundaries of the magnetic core loss model. This was achieved primarily through three approaches: 1) Introducing data augmentation strategies to endow the model with robustness in handling inputs with arbitrary phases and non-integer periods, thereby directly addressing non-ideal excitation waveforms encountered in practical engineering applications; 2) This research constructed

a more comprehensive experimental dataset by independently measuring data for three materials—nanocrystalline, iron powder, and ferrite—across a wide temperature range of -180°C to 120°C and a wide frequency band of 50 kHz to 300 kHz, including ferrite and nanocrystalline materials under magnetic saturation conditions, thereby filling gaps in existing datasets; 3) Systematic multi-scale validation was conducted. The model demonstrated excellent generalization capabilities on the extended dataset covering the aforementioned new materials and extreme operating conditions, and its accuracy was verified using publicly available ferrite databases. Collectively, these efforts have advanced core loss modeling from the pursuit of ideal accuracy to a practical stage that meets the demands of complex, real-world engineering scenarios.

The subsequent sections of this paper are organized as follows: Section II outlines the framework of the proposed modeling method, along with the implementation of data augmentation mechanisms and training optimization strategies. Section III details the data acquisition process based on the magnetic core loss testing platform and the model training procedure. Section IV introduces and discusses the model's prediction results for core losses in three materials. Section V describes the process of selecting ferrite as the material for further refined modeling and discusses the model prediction results. Finally, Section VI provides a summary of the entire paper.

II. FRAMEWORK OF THE PROPOSED MULTIMODAL FUSION NETWORKS BASED DATA-DRIVEN CORE LOSS MODEL

This section details the key components of the overall framework for the data-driven model, including the proposed data augmentation mechanism, training optimization methods, and modelling approach, while explaining its predictive mechanism.

A. Data Augmentation

Data augmentation and preprocessing lay a crucial foundation for enhancing the generalization capability and robustness of deep learning models in predicting magnetic core losses. In waveform data processing, augmentation techniques expand the effective training dataset by simulating variations in waveform patterns and system parameters. These techniques enable models to learn invariant features and reduce overfitting, proving particularly effective when datasets are limited in quantity or type.

The data augmentation methods employed in this study incorporate dynamic randomization operations, specifically: random periodic shifts, waveform flipping, noise injection, and parameter perturbation. Each technique carries specific physical significance: random shifts simulate phase variations in periodic excitation; flipping enhances symmetry recognition capabilities; noise injection improves tolerance to measurement errors; while parameter perturbation mimics natural fluctuations under operational conditions such as temperature and excitation amplitude variations. During training, these operations are applied randomly with a certain probability to significantly enhance the model's generalization capability towards unknown waveform characteristics. Fig. 1 shows the result of data augmentation ap-

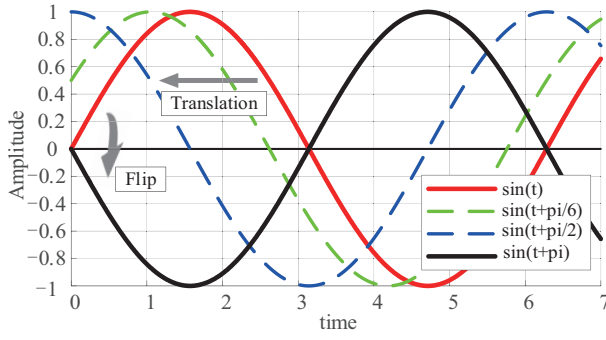


Fig. 1. Comparison of waveforms with different phase shifts and inversions.

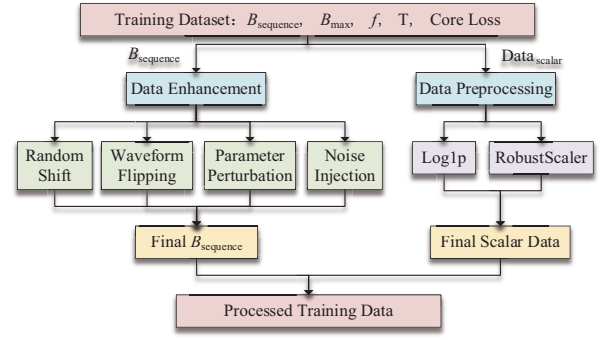


Fig. 2. Flowchart of data augmentation.

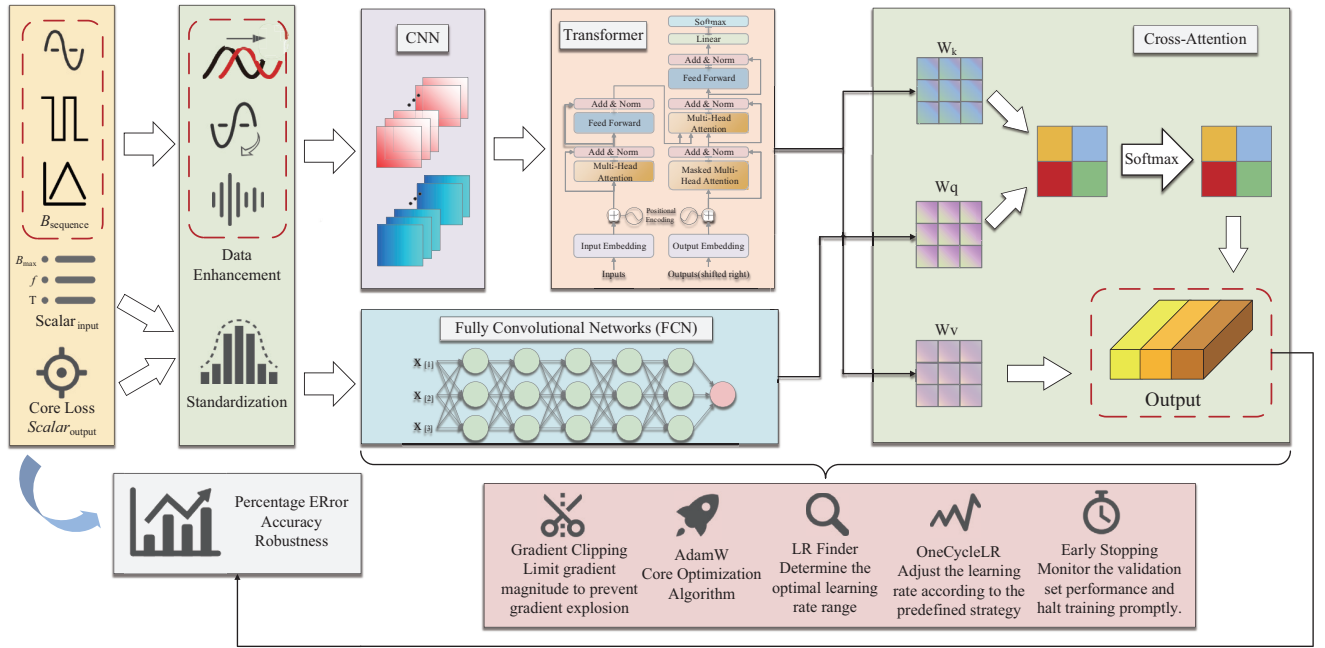


Fig. 3. Architecture of the multimodal fusion model based on data augmentation and training optimization strategies.

plied to a standard sinusoidal excitation waveform.

Given that core losses exhibit an approximately exponential relationship with influencing factors, standardizing MagNet data using conventional methods results in a right-skewed distribution, rendering model training susceptible to dominance by larger values. A dedicated preprocessing pipeline is implemented to normalize the input distribution and address the heavy-tailed nature of the data. This pipeline comprises:

Log1p: Training data values are transformed via the $\log(1+x)$ function. This transformation mitigates distribution skewness and compresses the dynamic range of target variables, thereby stabilizing gradient updates and enhancing convergence during training.

RobustScaler: Unlike standard scaling, RobustScaler scales features based on median and interquartile range. This ensures uniformly distributed standardized data, preventing common outliers in experimental magnetic data from adversely affecting model training [19].

The aforementioned data augmentation methods can simulate a wide range of physical scenarios while ensuring numerical

stability during optimization. Fig. 2 presents a flowchart of the data augmentation process, highlighting the application workflow for serialized transformation operations on input waveforms and parameters.

B. Training Optimization

Training optimization strategies play a crucial role in accelerating convergence speed and enhancing model accuracy and generalization capabilities. Within deep learning, complex mapping relationships are effectively approximated through parameter optimization. Consequently, implementing optimized training techniques constitutes an effective pathway to achieving stable and efficient model learning. These strategies hold significant value for accelerating training convergence, preventing overfitting, and improving model performance under diverse conditions. Fig. 3 illustrates the complete optimization framework and key components within this model.

The ReduceLROnPlateau learning rate strategy, combined with gradient clipping and early stopping mechanisms, forms

the core of this model's training optimization. Its implementation involves:

$$lr_{\text{new}} = \max(lr_{\text{old}} \times \text{factor}, lr_{\text{min}}) \quad (1)$$

where lr_{new} denotes the new learning rate after decay, lr_{old} represents the current learning rate, factor signifies the attenuation multiplier, and the lr_{min} indicates the set lower limit.

Gradient clipping refers to proportionally scaling down the entire gradient vector based on a preset threshold during training to prevent gradient explosion. Early stopping refers to prematurely terminating training based on the trend of validation set performance during training to prevent model overfitting. The LR Finder primarily determines an appropriate learning rate range during the early stages of training by incrementally increasing the learning rate and observing changes in the loss.

AdamW represents a significant refinement of the traditional optimizer Adam, integrating momentum with adaptive learning rates [20]. It dynamically adjusts the learning rate for each parameter by computing the first and second moments of the gradient. Compared to Adam, AdamW decouples weight decay from gradient computation, thereby mitigating the impact of gradient fluctuations. This renders it more efficient when training large-scale models such as Transformers.

C. CNN-Transformer-Attention Neural Network

In deep learning research, deploying a single model inherently possesses both strengths and limitations. Combining multiple models effectively leverages the advantages of each while compensating for the shortcomings of any single model, thereby enhancing training accuracy.

This paper employs a CNN-transformer-attention neural network architecture, comprising a total of 86,385 trainable parameters. Its detailed architecture is described as follows: The CNN module consists of four one-dimensional convolutional layers and two pooling layers, serving as the backbone for local feature extraction. This module processes raw waveform inputs, with each sample containing 1024 data points. Through alternating convolution, pooling operations, and residual connections, the sequence length is progressively compressed to 32 while increasing the number of feature channels to the preset embedding dimension. During this process, the convolution layers capture multi-scale local temporal patterns, while the pooling layers expand the receptive field through down sampling and enhance translation invariance. Each convolutional layer is followed by a batch normalization layer and a GELU activation function.

The transformer encoder consists of three stacked encoder blocks. Each encoder block contains a self-attention mechanism equipped with four attention heads and a feedforward network. The self-attention mechanism enables all information within the sequence to mutually attend to each other, calculating their relative importance. The design of multiple attention heads allows the model to focus on different subspaces of information in parallel. Layer normalization and residual connections are applied around each sublayer to ensure training stability and

gradient flow.

The parameter branch consists of a single fully connected layer designed to process auxiliary physical parameters. This branch projects six input parameters—including the log-transformed peak magnetic flux density, frequency, temperature, and three waveform shape descriptors—into an embedding dimension space identical to the waveform features. The data then undergoes processing through a GELU activation function and a dropout layer.

The Attention is an independent fusion module that performs conditional feature selection. Within this module, the embedding vector output from the parameter branch serves as the query, while the 32 waveform feature sequences output from the CNN-Transformer path act as both keys and values. This mechanism enables the model to dynamically select and weight the most relevant temporal features within the waveform based on specific operational parameters.

The fusion network consists of two fully connected layers for generating the final prediction. It concatenates features filtered through cross-attention with globally pooled features from the waveform sequence, then sequentially passes them through two linear transformation layers. The final output is a scalar value representing the predicted core loss. GELU activation and dropout operations are also applied between the two layers.

In summary, as shown in Fig. 3, the preprocessed data is fed in parallel to the convolutional neural network-transformer pathway and the parameter branch. The Convolutional Neural Network extracts multi-scale local patterns, while the Transformer captures global sequence dependencies, jointly forming complementary feature representations. Subsequently, the cross-attention mechanism enables deep, conditionally dependent fusion of these multimodal features. The fused representations are ultimately processed by the subsequent network to achieve end-to-end core loss prediction.

III. DATA-DRIVEN TRAINING AND OPTIMIZATION

To validate the effectiveness of the multimodal neural network model, the network architecture shown in Fig. 3 was constructed. All relevant algorithms in this paper were implemented using Python and the PyTorch library. This section will detail the data acquisition process and the model training and optimization procedures.

A. Data Collection

In high-frequency applications, magnetic materials typically require high saturation flux density, high permeability, and low core loss. Currently, the most commonly used core material is ferrite, which offers advantages such as high permeability, low core loss, high Curie temperature, and low cost, but its saturation flux density is relatively low. Iron powder cores offer high saturation flux density and excellent temperature stability, but they have lower permeability and higher high-frequency losses. Nanocrystalline materials exhibit outstanding magnetic properties, high permeability, and low core losses, but they come with relatively higher costs. This study selected ferrite cores of the Ferrite M77 type, nanocrystalline cores of the

TABLE I
PARAMETERS OF THE MAGNETIC CORES

Parameters	M77	TFN-CM	T200-52B
Material type	Ferrite	Nanocrystalline	Iron powder
Manufacturer	Fair-Rite	Pourleroi	Micrometals
μ_r	2000	80000	75
Saturation flux density	510 mT	1228 mT	1400 mT
L_e	6.17 cm	14.07 cm	13 cm
A_e	0.39 cm ²	0.78 cm ²	2.32 cm ²
Magnetic saturation	Yes	Yes	No
Temperature		-180 °C ~ 120 °C	
Frequency		50 kHz ~ 300 kHz	

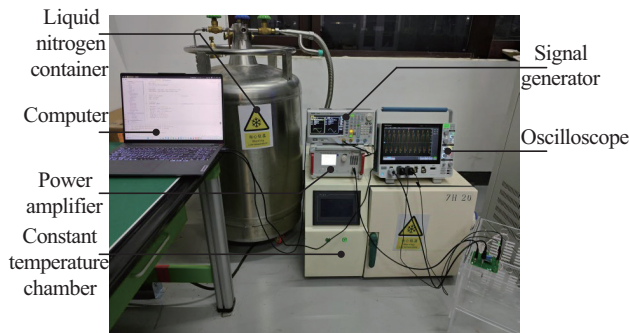


Fig. 4. Platform for magnetic core loss test.

Nanocrystalline TFN-CM type, and iron powder cores of the Iron powder T200-52B type as research subjects, with their specific parameters shown in Table I.

The temperature range tested in this study spans -180 °C to 120 °C, with a frequency range of 50 kHz to 300 kHz. Both Ferrite M77 and Nanocrystalline TFN-CM materials exhibited magnetic saturation conditions, and the corresponding saturation data were incorporated into the training dataset. The Iron powder T200-52B material exhibited extremely low magnetization inductance values due to its low permeability. Combined with the limited output capability of the test platform's power amplifier, magnetic saturation conditions could not be observed.

To minimize errors from manual measurements, an automated measurement method based on the pyvisa library was employed to collect large-scale, high-quality data. The experimental circuit for measuring magnetic core losses was constructed using the dual-winding method described in [21], which involves connecting a resonant capacitor in series with the excitation winding. The test platform is shown in Fig. 4. The core under test was placed in a constant temperature chamber to maintain a constant ambient temperature. A computer simultaneously controlled the signal generator and oscilloscope to perform automated testing. After testing, Python was used to clean the data to ensure its quality.

The following presents measurement data using Ferrite M77 material as an example. Fig. 5 shows the trend of saturation flux density versus temperature for Ferrite M77. Fig. 6 illustrates

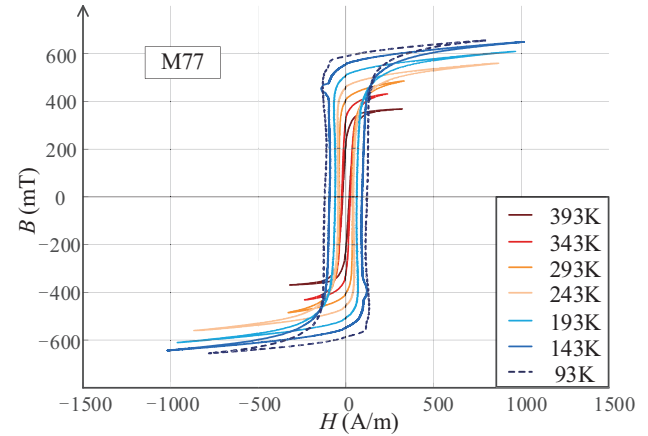


Fig. 5. Trend of the saturation flux density of M77 material as a function of temperature.

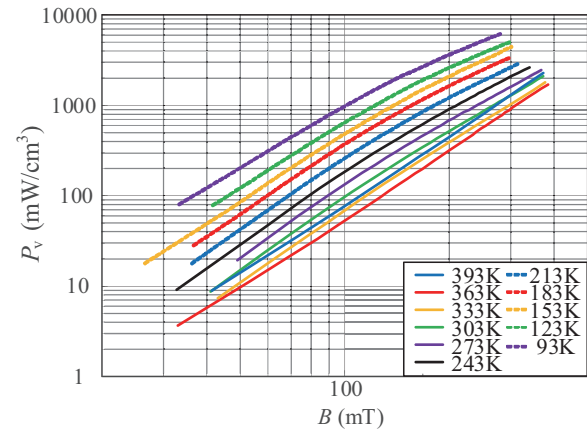


Fig. 6. Variation of Core Loss in Ferrite M77 Material at 200 kHz Frequency with Temperature and Magnetic Induction Strength.

the relationship between core loss and temperature/magnetic flux density for Ferrite M77 material. The test results exhibit smooth and flat curves consistent with physical principles, demonstrating the accuracy of this measurement method.

Compared with public databases such as MagNet, which contain only ferrite materials, this study independently obtained additional loss data for nanocrystalline and iron powder cores through experimental testing. Furthermore, the temperature range was expanded to a wide range of -180 °C to 120 °C, and the testing range for both ferrite and nanocrystalline materials was extended to the state of magnetic saturation. This lays a critical data foundation for constructing more universally applicable models.

B. Data Processing

As previously stated, the initial phase of the magnetic flux density waveform within a single cycle should not affect the final core loss. However, the initial phase of the magnetic flux density waveform obtained is typically set to zero. This makes it difficult for the model to adapt to excitation waveforms under different conditions during training. Therefore, the

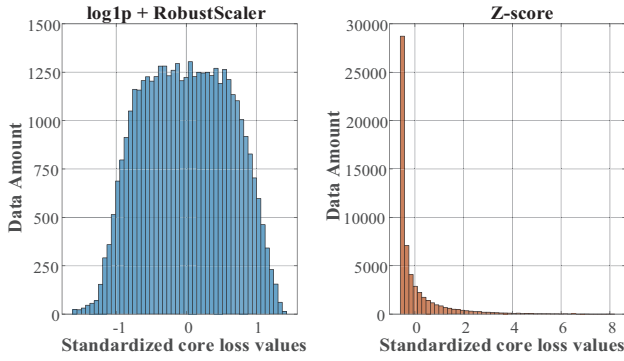


Fig. 7. Comparison of two standardized datasets.

model incorporates 30% amplitude-phase random offsets, 50% probability waveform reversals, and a certain proportion of parameter perturbations into the magnetic induction intensity waveform. The newly generated data is then integrated into the original dataset for joint training, enhancing physical interpretability and adaptability when encountering diverse input conditions during training.

In the raw data used for model training, inconsistent units result in differing data scales. Direct training under such conditions leads to a sharp decline in training efficiency and severe distortion of gradients. Therefore, standardization is commonly employed to eliminate scale differences. However, as previously mentioned, since core losses and influencing factors exhibit an approximately exponential relationship, conventional standardization methods cause the data center of gravity to shift. Consequently, this model employs Log1p and RobustScaler for standardization.

As shown in Fig. 7, this heatmap illustrates the distribution of data processed using the Log1p and RobustScaler method and the traditional normalization method, using magnetic core loss data as an example. It is evident that the data processed with the Log1p and RobustScaler normalization method (left panel) exhibits a more uniform distribution within the calibration range. In contrast, the data processed with the traditional normalization method (right panel) retains the distribution characteristics of the original data, showing a more extreme distribution pattern.

C. Model Parameter Tuning and Training Optimization

During the training process of data-driven models, selecting appropriate hyperparameters is one of the key factors affecting performance [22]. In this paper, relevant hyperparameters include training configuration parameters, parameters for neural network architecture at various levels, and LR scheduling parameters. All these parameters require manual presetting.

This study employs mean error, 95th percentile error, and maximum error as metrics for model accuracy. Mean error represents the average deviation across all model predictions, reflecting overall precision. The 95th percentile error represents the error range in the vast majority of cases, making it highly representative. The maximum error reveals the model's performance under worst-case conditions. Its large value may stem from extreme conditions never covered in the training data, or

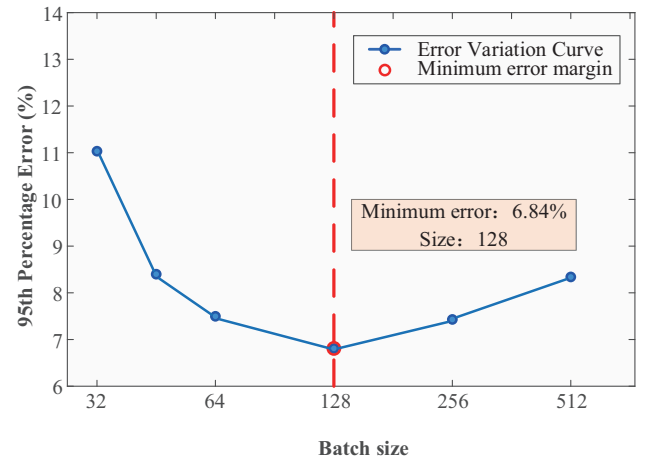


Fig. 8. Trend of 95th percentage error with Batch size.

from small true values resulting in a small denominator when calculating the percentage error, leading to a large overall value. Therefore, metrics such as the mean error and 95th percentile error should be prioritized for reference. In this paper, the errors mentioned above were calculated using the validation set, and the data in the validation set were randomly sampled from the original dataset at a specific ratio.

Training configuration parameters primarily include batch size, weight decay coefficient, gradient clipping threshold, and maximum training epochs. Taking batch size as an example, setting it too small can lead to unstable training, slow convergence, and excessive training time; conversely, setting it too large may cause overfitting, impair final generalization performance, and increase computational resource requirements. Therefore, these parameters require fine-tuning based on hyperparameter optimization experience and gradient trends. Fig. 8 illustrates the model accuracy comparison across different configurations at various batch sizes.

Model architecture parameters primarily include the number of layers in each neural network level, core feature dimensions, the number of attention heads, and dropout rates at each level. The learning rate scheduling parameters include the initial learning rate factor and the final learning rate factor within linear warmup. These parameters were obtained through the aforementioned methods and are presented in Table II.

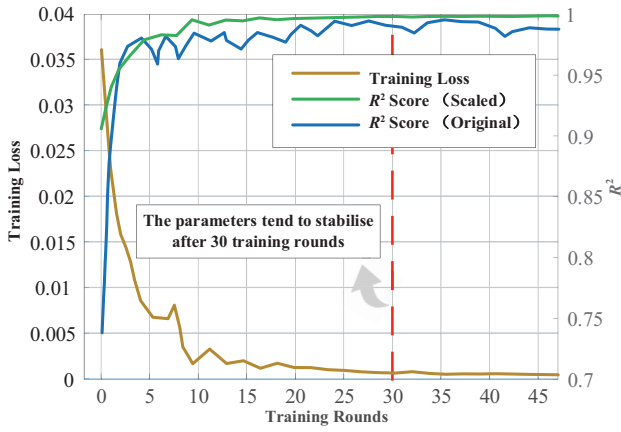
Through the above adjustments, the optimal model was obtained. To visually demonstrate the model's convergence process, Fig. 9 displays the training loss and R^2 value for each training cycle. It can be observed that the training loss decreases rapidly between rounds 0 and 10, stabilizing after round 30 and approaching 0.001. The R-squared value rises sharply during the initial training phase before stabilizing, ultimately reaching 0.9991. This fully demonstrates the high efficiency and accuracy of this model. All models in this study were trained on a single NVIDIA RTX 5090 graphics card.

IV. RESULTS VERIFICATION AND DISCUSSION

After completing model training, this research begin analyzing the model's prediction results. This chapter focuses on four

TABLE II
FINAL MODEL PARAMETERS

Parameters	Value	Definition
embed_size	48	Core feature dimensions
num_heads	4	Number of Attention points
dropout	0.1	Dropout rate, preventing overfitting
augment_prob	0.5	Data augmentation probability
batch_size	64~512	Training batch size
weight_decay	1e-4	Weight decay coefficient
max_epochs	500	Maximum training rounds
start_factor	0.1	Initial learning rate factor
patience	100	Number of waiting cycles
end_factor	1.0	Final learning rate factor
num_iter	100	Number of search iterations

Fig. 9. Trends of loss and R^2 with training rounds.

main areas: an overview of the training results, a comparison of different models, data augmentation methods, and predictions for extreme operating conditions.

A. Analysis of Results

As shown in Fig. 10, the models for the three materials—Ferrite M77, Iron powder T200-52B and Nanocrystalline TFN-CM—all demonstrated excellent learning performance and convergence characteristics during training. The training loss initially decreased rapidly with increasing training steps, stabilizing around the 20th iteration and ultimately settling at an extremely low level. Concurrently, the R^2 value rose sharply and approached 1, indicating the model's continuously enhanced ability to capture and fit data features.

Although the loss function and R^2 are commonly used to characterize model accuracy during training, they often lack tangible physical meaning based solely on the data itself. Therefore, this paper evaluates model accuracy by comparing the errors between predicted and actual values, such as the mean percentage error, 50th percentile error, 95th percentile error, and maximum error, as assessment criteria [23], [24]. The percentile error calculation for core losses is as follows:

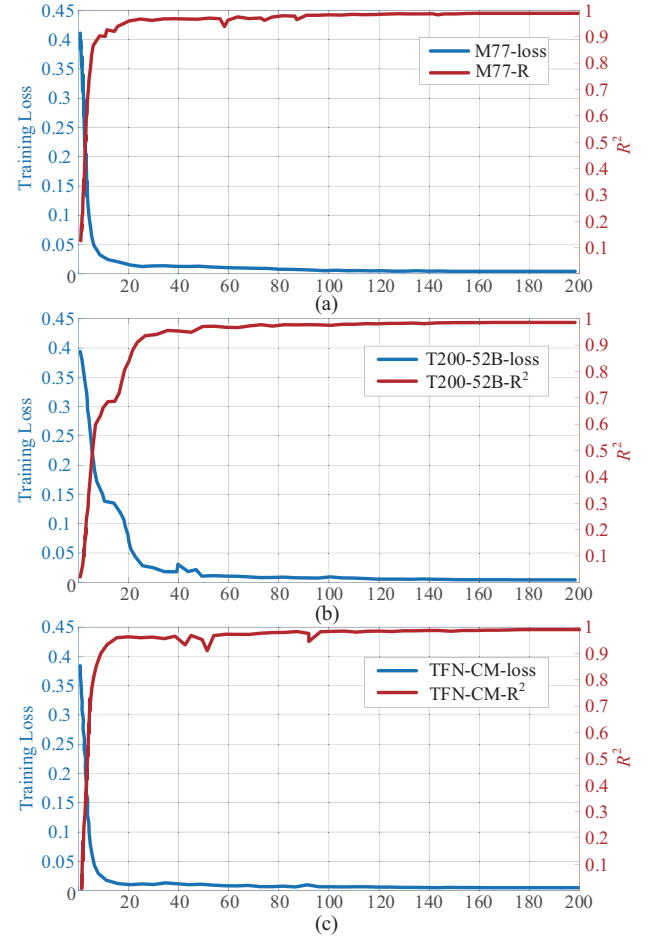
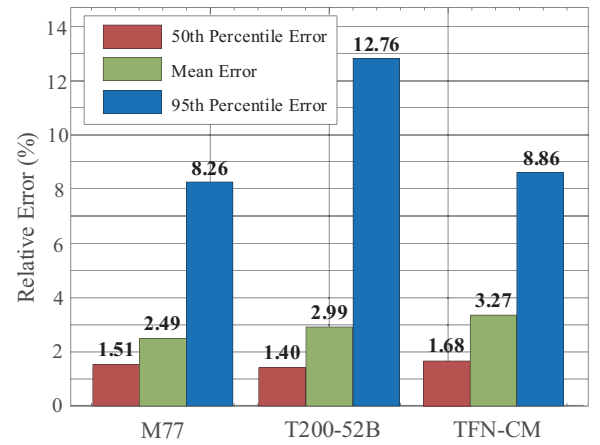
Fig. 10. Loss and R^2 for training with three materials. (a) M77. (b) T200-52B. (c) TFN-CM.

Fig. 11. Relative error under three material predictions.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{i_real} - y_{i_predict}}{y_{i_real}} \right| \times 100\% \quad (2)$$

Fig. 11 clearly demonstrates that the models established for the three different materials achieve excellent predictive accuracy and robustness. Overall, the models maintain low prediction errors for the vast majority of samples, with the 50th

TABLE III
PERFORMANCE COMPARISON AMONG DIFFERENT MODELS

Model	Mean error		
	M77	TFN-CM	T200-52B
CNN	57.85%	46.85%	50.54 %
Transformer	43.6 %	25.62 %	13.46 %
CNN-transformer	8.15 %	10.26 %	4.81 %
This model	2.49 %	2.99 %	3.27 %

percentile error consistently below 2% and the 95th percentile error consistently below 15%. This indicates that the models exhibit outstanding robustness and reliability across different materials

B. Model Comparison

To further validate the effectiveness of the proposed model architecture, ablation experiments were conducted to compare three model variants—CNN, Transformer, CNN-Transformer, and the CNN-transformer-attention model introduced herein—and evaluate their predictive performance across the three core materials. The configuration parameters of all model variants were identical to those of the original model, with all other conditions held constant. As shown in Table III, this model achieved the highest accuracy across all three materials. This fully confirms the contribution of each component to the model's performance.

C. Analysis of Data Augmentation

As mentioned earlier, the initial phase of the magnetic flux density waveform should be unaffected by core losses. To verify the model's accuracy in this regard, the raw data was adjusted using a phase random shift of 50% via data processing code and saved as a new dataset. The model was then compared with a model that underwent no data augmentation but shared identical parameters for prediction. The results are shown in Fig. 13 and Fig. 14.

Fig. 12 and Fig. 13 respectively illustrate the discrepancy between predicted and actual values for the two models, along with their correlation coefficients. It is evident that the model employing data augmentation in Fig. 13 demonstrates superior adaptability and accuracy across magnetic induction waveforms with varying initial phases. Its predicted values nearly perfectly align with actual values, achieving an R^2 value of 0.9770. In contrast, the model without data augmentation in Fig. 14 achieves an R^2 of only 0.6926, with noticeable deviations between predicted and actual values.

Additionally, this model employs data augmentation techniques to enable predictions based on sequences of varying lengths. Taking the 3C90 material model as an example, Fig. 14 illustrates the average percentage error of the model's predictions when input sequences span 150% of the cycle, 125% of the cycle, 100% of the cycle, 75% of the cycle, and 25% of the cycle. The results indicate that when the input sequence

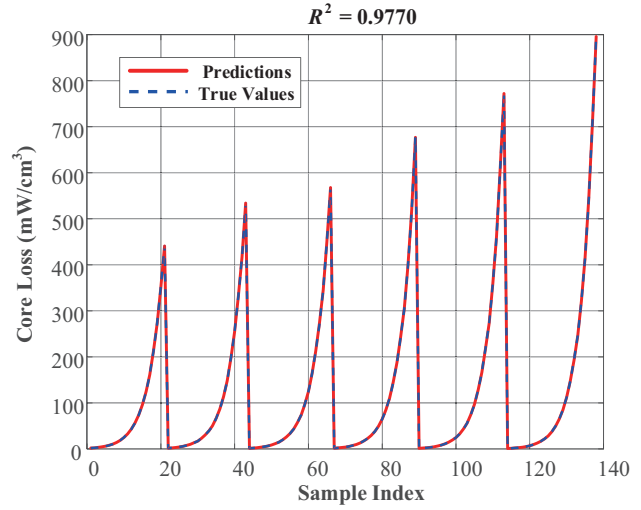


Fig. 12. Comparison of predicted values to actual values after applying data augmentation.

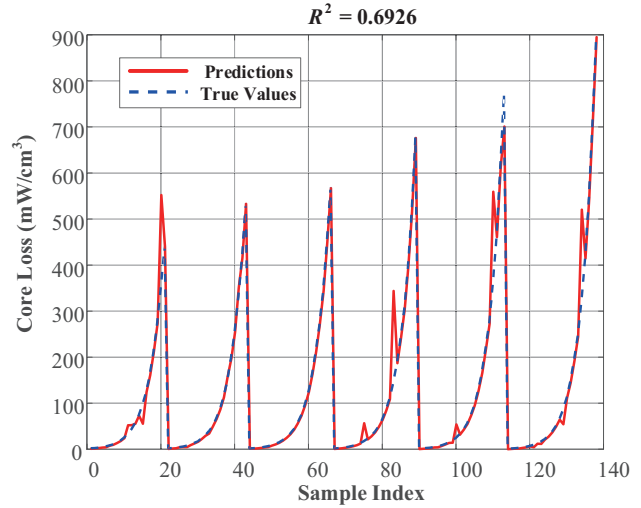


Fig. 13. Comparison of predicted values and actual values without data augmentation.

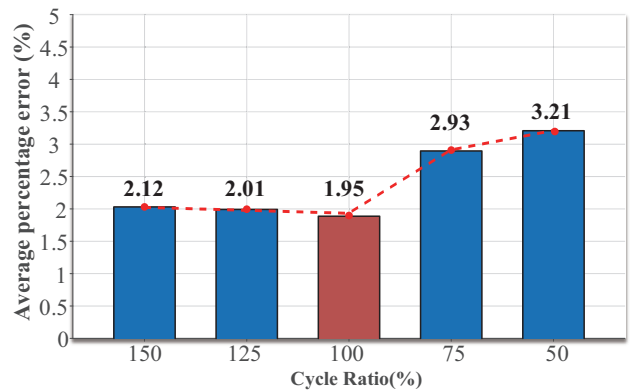


Fig. 14. Trend of percentage error with sequence length.

length exceeds one cycle, prediction accuracy remains virtually unaffected. When the sequence is shorter than one cycle, accu-

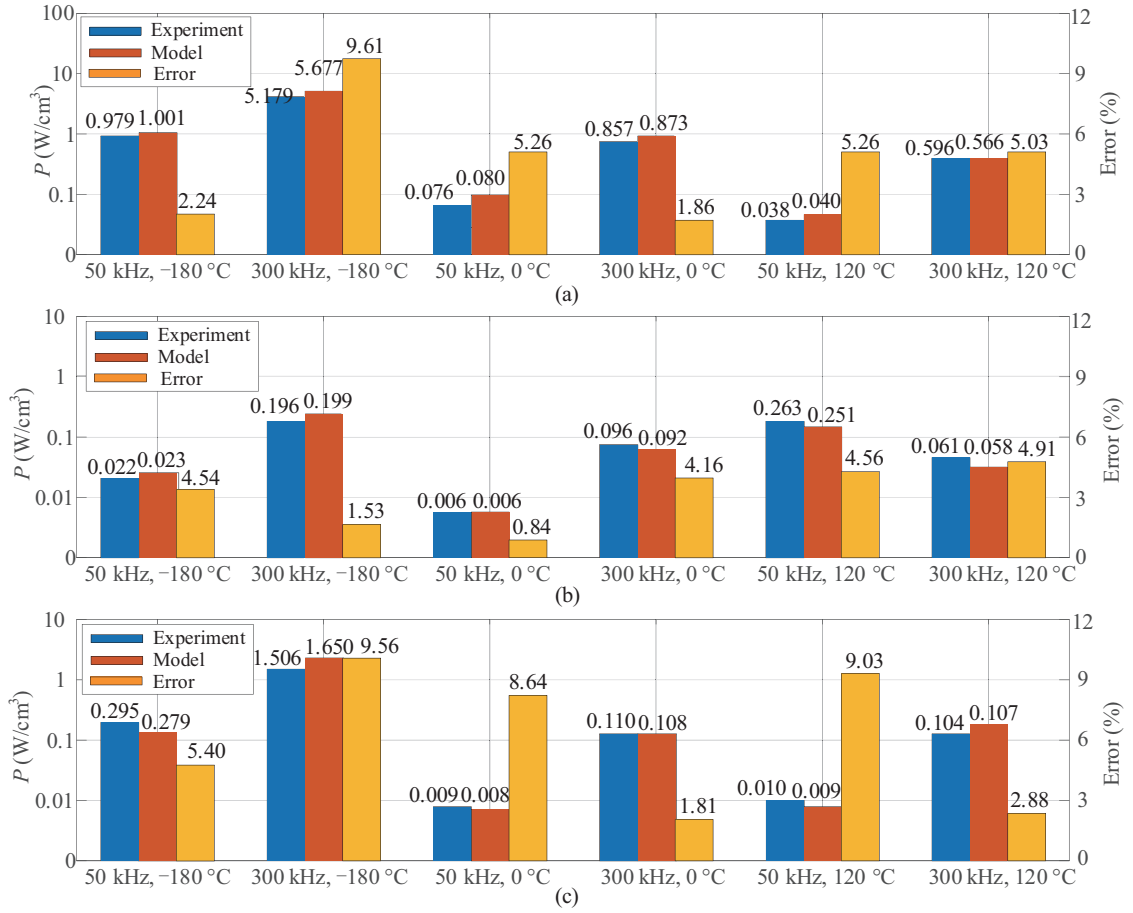


Fig. 15. Comparison of predicted results and actual values for three materials under extreme frequencies and temperatures. (a) M77-100mT. (b) TFN-CM-50mT. (c) T200-52B-10mT.

racy slightly decreases due to incomplete waveforms, yet the percentage error still remains below 12%. This demonstrates the model's robust capability to capture temporal characteristics. These results demonstrate that the data augmentation optimization strategy significantly enhances both the physical interpretability and accuracy of the model.

D. Analysis of Extreme Operating Conditions

Fig. 15 shows the model's predictions for the three core materials under extreme operating conditions after incorporating data from extreme conditions such as low temperature, high temperature, and high frequency into the training set. Overall, the model maintains high accuracy for all three materials and shows stable performance across the covered extreme operating conditions.

Fig. 16 and Fig. 17 show a comparison of the model's predictions with actual values at different frequencies at 20 °C, at which point both the ferrite M77 and the nanocrystalline TFN-CM material have reached magnetic saturation. Magnetic saturation data for both materials were included in the training set for model training. As shown in the figure, the model's predicted values for both materials closely match the actual values across the entire frequency range, with small absolute errors, demonstrating the model's excellent predictive accuracy under magnetic saturation conditions.

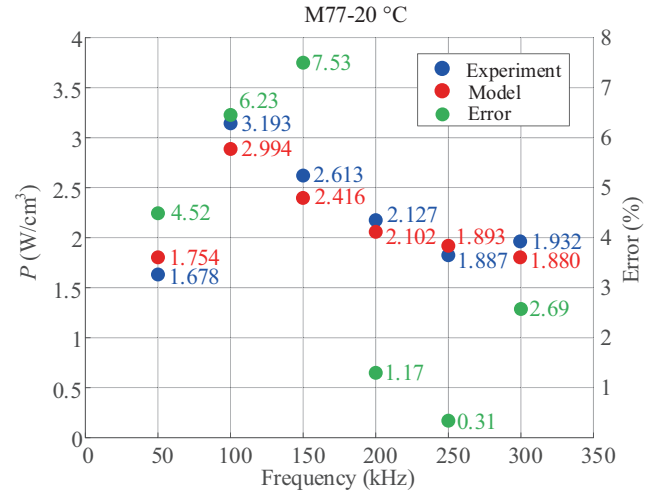


Fig. 16. Comparison of predicted results and actual values for Ferrite M77 material at magnetic saturation.

V. MODELING OF MULTIPLE FERRITE CORE LOSSES

This section primarily addresses loss modeling for various core types using ferrite as the specific material. The content covers data selection, analysis of training results, and accuracy comparisons.

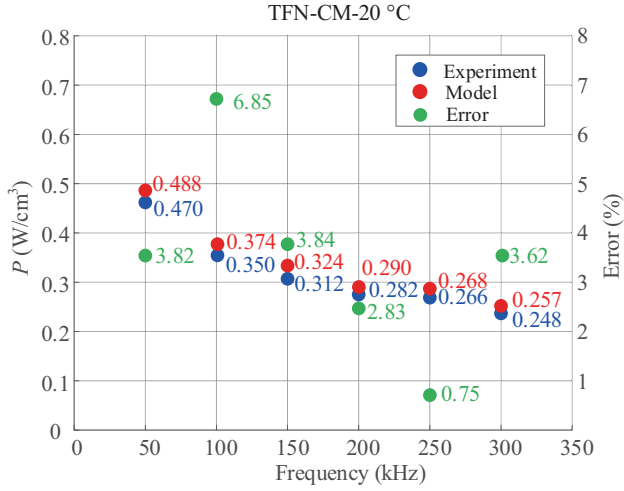


Fig. 17. Comparison of predicted results and actual values for Nanocrystalline TFN-CM material at magnetic saturation.

A. Data Selection

This paper employs the database provided by the MagNet Challenge 2023 for model training. This database utilizes a dual-winding method to measure core losses and employs computerized automatic control for data acquisition. This database covers a wide range of ferrite materials manufactured by Ferroxcube and TDK. Operating conditions include waveforms such as sine waves, triangular waves, and square waves with arbitrary duty cycles, spanning a frequency range from 50 kHz to 500 kHz and a temperature range from 25 °C to 75 °C. For a detailed description of the MagNet database acquisition system, please refer to [25]–[27]. This section uses the 3C94 material as an example to explain the model training and optimization process in detail. The remaining materials are employed for in-depth analysis of the model's accuracy and robustness.

B. Training Results

As shown in Fig. 18, the graph displays the magnetic loss error output after model training using data from nine core materials in the MagNet database as training data. It is evident that the data-driven model proposed in this paper demonstrates outstanding performance in predicting core losses across different materials. The average prediction error for most materials remains below 3.4%, and the maximum error for all materials is less than 27%. The nine models trained on the nine material datasets respectively maintain high accuracy on their respective materials.

Fig. 19 illustrates the average prediction error of the model for the three materials—N87, 3C94, and 3E6—under sine, triangular, and square wave conditions, respectively. The figure shows that the model maintains a low average prediction error for all three materials across all excitation waveforms, demonstrating the model's good adaptability to different waveforms.

To further quantify the model's efficiency, Table IV summarizes the training dataset size, number of model parameters, training duration, and response time for each material dataset.

The response time refers to the time required to predict 5,000 new data points using the trained model. The results demonstrate that the proposed model achieves high accuracy and efficiency while maintaining low training costs.

C. Loss Prediction Under Different Methods

Existing literature has proposed various methods for predicting magnetic core losses, such as the Modified Steinmetz Equation (GSE and iGSE), LSTM + DNN, the Corrected Steinmetz Random Forest Model (Hybrid Model), the Physically Mechanism-Driven Model, and the Vision Transformers Model (MAG-ViT). This paper utilizes data from N87 magnetic core material in the MagNet database as the training dataset. The mean percentage error is selected as the evaluation metric, with testing conducted on multiple waveforms. The model was trained in parallel with the aforementioned models under identical dataset conditions to compare their accuracy and material adaptability.

Table V presents a comparison of existing literature in this research field using the N87 magnetic core material model as an example. The results indicate that, although the model proposed in this study did not achieve the best performance among all existing models, its accuracy and parameter size remain at a relatively high level. This study extends the model's functionality by incorporating data augmentation techniques such as arbitrary phase and sequence length inputs, and no other references in this field has addressed this work. Furthermore, it does not solely rely on ferrite material data from the MagNet database. Instead, experimental data from ferrite powder cores and nanocrystalline materials were obtained through a physically constructed test platform and used for model training. This demonstrates the magnetic loss prediction model's robust performance across diverse core materials. Simultaneously, data from Nanocrystalline TFN-CM and Ferrite M77 materials under magnetic saturation conditions were collected and incorporated into the training set, whereas the models in other studies were trained on datasets that did not include data on multi-material or extreme operating conditions.

The aforementioned work represents areas not previously addressed in existing literature. This study marks the first systematic integration of support for all these critical characteristics in practical engineering applications. Therefore, through comprehensive comparison, the model proposed in this paper holds significant importance for modeling iron core losses in engineering applications.

VI. CONCLUSION

This study proposes a comprehensive core loss modeling framework encompassing specialized data augmentation strategies, multimodal neural network architectures, and multiscale validation paradigms. By integrating CNN's local feature extraction capabilities, Transformers' global sequence modeling advantages, and multimodal information fusion through attention mechanisms, it effectively overcomes the limitations of traditional Steinmetz equations in complex waveforms and

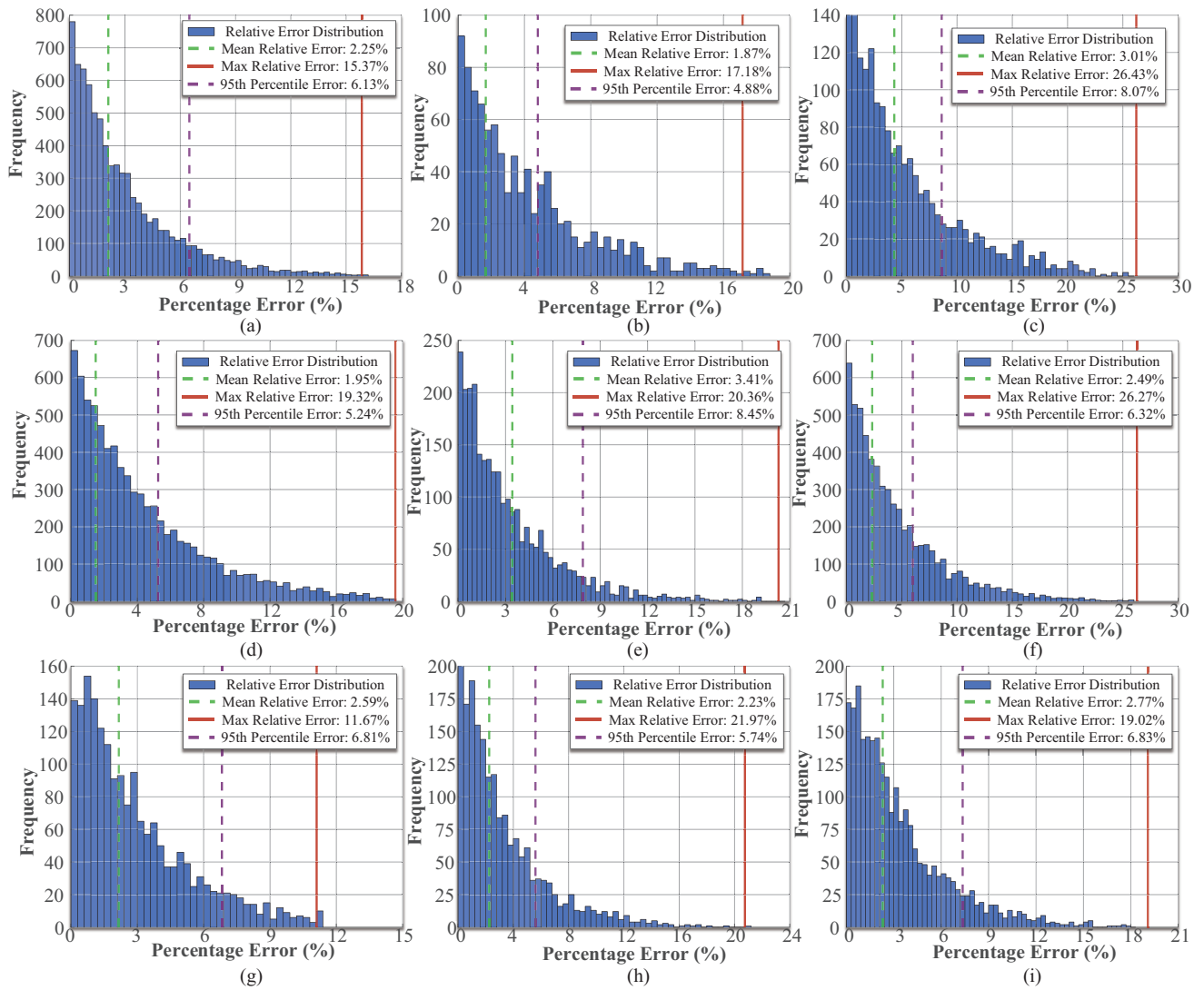


Fig. 18. Histogram of core loss error ratios for nine different material groups after training with given data. (a) N87. (b) 3E6. (c) N49. (d) 3C94. (e) 3F4. (f) 3C90. (g) N27. (h) N30. (i) 78.



Fig. 19. Mean error of model predictions under three waveforms.

TABLE IV
MODEL COMPARISON ACROSS DIFFERENT LITERATURE

Material	Training data size	Training time	Response time	Model size
M77	37.2 MB	933 s	0.66 s	86385 Parameters
T200-52B	39.2 MB	660 s	0.69 s	
TFN-CM	45.9 MB	963 s	0.74 s	
3C90	493 MB	893 s	0.94 s	
3C94	373 MB	818 s	0.84 s	
3E6	84.8 MB	301 s	0.79 s	
3F4	79.5 MB	342 s	1.01 s	
N27	138 MB	747 s	0.88 s	384 KB
N30	108 MB	457 s	0.94 s	
N49	104 MB	426 s	0.73 s	
N87	492 MB	886 s	1.03 s	
77	138 MB	782 s	0.97 s	

Note: All of the parameters were obtained by RTX 5090.

high-frequency scenarios. Simultaneously, the study integrates extensive experimental data—including novel materials and saturated operating conditions—with physics-based data augmentation techniques. This significantly enhances the model's

TABLE IV
THE TRAINED MODEL INFORMATION FOR DIFFERENT MATERIALS

Reference	Error	Parameters	Model	Multiple waveforms	Arbitrary phase input	Non-integer period input	Magnetic saturation	Multiple materials
[28]	29.13%	3	GSE	Unknown	Unknown	Unknown	Not covered	Not covered
[28]	24.93%	3	iGSE	Unknown	Unknown	Unknown	Not covered	Not covered
[26]	4.53%	1033729	LSTM+DNN	Yes	Unknown	Unknown	Not covered	Not covered
[28]	6.41%	11012900	Hybrid model	Yes	Unknown	Unknown	Not covered	Not covered
[14]	2.30%	139938	Physics-inspired	Yes	Unknown	Unknown	Not covered	Not covered
[26]	1.58%	28564	MAG-ViT	Yes	Unknown	Unknown	Not covered	Not covered
This paper	2.25%	86385	CNN-transformer - attention	Yes	Yes	Yes	Yes	Yes

generalization capability and applicability, achieving more comprehensive functional coverage. It provides a practical solution for efficiently modeling core losses in power electronic systems.

REFERENCES

- [1] W. Roshen, "Ferrite core loss for power magnetic components design," in *IEEE Transactions on Magnetics*, vol. 27, no. 6, pp. 4407–4415, Nov. 1991.
- [2] A. J. Hanson, J. A. Belk, S. Lim, C. R. Sullivan, and D. J. Perreault, "Measurements and performance factor comparisons of magnetic materials at high frequency," in *IEEE Transactions on Power Electronics*, vol. 31, no. 11, pp. 7909–7925, Nov. 2016.
- [3] K. Venkatachalam, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," in *2002 IEEE Workshop on Computers in Power Electronics, 2002. Proceedings.*, Mayaguez, PR, USA, 2002, pp. 36–41.
- [4] H. Li, D. Serrano, S. Wang, and M. Chen, "MagNet-AI: neural network as datasheet for magnetics modeling and material recommendation," in *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15854–15869, Dec. 2023.
- [5] J. Reinert, A. Brockmeyer, and R. W. A. A. De Doncker, "Calculation of losses in ferro- and ferrimagnetic materials based on the modified Steinmetz equation," in *IEEE Transactions on Industry Applications*, vol. 37, no. 4, pp. 1055–1061, Jul.-Aug. 2001.
- [6] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," in *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 964–973, Feb. 2012.
- [7] Y. Han, G. Cheung, A. Li, C. R. Sullivan, and D. J. Perreault, "Evaluation of magnetic materials for very high frequency power applications," in *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 425–435, Jan. 2012.
- [8] B. Su, K. Sun, M. Yang, Y. Xiao, K. Zhang, and B. Liu, "A core loss estimation method based on data-driven technology with multi-head attention mechanism," in *2024 IEEE 10th International Power Electronics and Motion Control Conference (IPEMC2024-ECCE Asia)*, Chengdu, China, 2024, pp. 3992–3996.
- [9] H. Wang, J. Kuang, and J. Shao, "Data-driven core loss modeling of magnetics," in *2025 IEEE 3rd International Conference on Control, Electronics and Computer Technology (ICCECT)*, Jilin, China, 2025, pp. 1428–1432.
- [10] X. Ma, "APOR: a data-driven framework for accurate magnetic core loss prediction," in *2025 2nd International Conference on Telecommunications and Power Electronics (TELEPE)*, Frankfurt, Germany, 2025, pp. 480–484.
- [11] J. Chen, Y. Liang, X. Li, X. Liu, H. Jiang, X. Miao, and L. Zhang, "A magnetic core loss modeling method based on domain adaptation using multiple kernel maximum mean Discrepancy," in *IEEE Transactions on Power Electronics*, vol. 41, no. 4, pp. 5103–5114, Apr. 2026.
- [12] X. Shen, H. Wouters, and W. Martinez, "Deep neural network for magnetic core loss estimation using the MagNet experimental database," in *2022 24th European Conference on Power Electronics and Applications (EPE'22 ECCE Europe)*, Hanover, Germany, 2022, pp. 1–8.
- [13] J. Deng, W. Wang, Z. Ning, P. Venugopal, J. Popovic, and G. Rietveld, "High-frequency core loss modeling based on knowledge-aware artificial neural network," in *IEEE Transactions on Power Electronics*, vol. 39, no. 2, pp. 1968–1973, Feb. 2024.
- [14] Y. Hu, J. Xu, J. Wang, and W. Xu, "Physics-inspired multimodal feature fusion cascaded networks for data-driven magnetic core loss modeling," in *IEEE Transactions on Power Electronics*, vol. 39, no. 9, pp. 11356–11367, Sept. 2024.
- [15] T. Chen, B. Zhang, T. Dong, S. Ma, and M. Zhang, "Characteristics of losses in composite material transformer cores," in *2024 IEEE International Magnetic Conference - Short papers (INTERMAG Short papers)*, Rio de Janeiro, Brazil, 2024, pp. 1–2.
- [16] H. Li, D. Yuan, S. Wang, S. Wang, and J. Zhu, "Core losses calculation of nanocrystalline alloy high frequency transformer considering magnetic hysteresis effects," in *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, Sydney, NSW, Australia, 2017, pp. 1–4.
- [17] P. Marketos, J. P. Hall, and S. E. Zirka, "Power loss measurement and prediction of soft magnetic powder composites magnetized under sinusoidal and nonsinusoidal excitation," in *IEEE Transactions on Magnetics*, vol. 44, no. 11, pp. 3847–3850, Nov. 2008.
- [18] A. Hilal, M.-A. Raulet, C. Martin, and F. Sixdenier, "Power loss prediction and precise modeling of magnetic powder components in DC-DC power converter application," in *IEEE Transactions on Power Electronics*, vol. 30, no. 4, pp. 2232–2238, Apr. 2015.
- [19] H. Qian, Q. Wen, L. Sun, J. Gu, Q. Niu, and Z. Tang, "RobustScaler: QoS-Aware Autoscaling for Complex Workloads," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, IEEE, pp. 2762–2775, May 2022.
- [20] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards understanding convergence and generalization of AdamW," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 9, pp. 6486–6493, Sept. 2024.
- [21] M. Mu, "High frequency magnetic core loss study," Ph.D. dissertation, Dept. Elect. Eng., Virginia Tech, Blacksburg, VA, USA, 2013.
- [22] C. Huang, Y. Li, and X. Yao, "A survey of automatic parameter tuning methods for metaheuristics," in *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 2, pp. 201–216, Apr. 2020.
- [23] Y. Setiyadi, J. Santoso, and K. Surendro, "Comparison of loss function weighting methods for multi-loss function GAN," in *2024 Ninth International Conference on Informatics and Computing (ICIC)*, Medan, Indonesia, 2024, pp. 1–6.
- [24] J. Xie, L. Song, and H. Huang, "Thermal radiation bias correction for infrared images using huber function-based loss," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
- [25] D. Serrano, H. Li, S. Wang, T. Guillod, M. Luo, V. Bansal, N. K. Jha,

Y. Chen, C. R. Sullivan, and M. Chen, “Why MagNet: quantifying the complexity of modeling power magnetic material characteristics,” in *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14292–14316, Nov. 2023.

- [26] M. Chen, H. Li, S. Wang, T. Guillod, D. Serrano, N. Förster, W. Kirchgässner, T. Piepenbrock, O. Schweins, O. Wallscheid, *et al.*, “MagNet challenge for data-driven power magnetics modeling,” in *IEEE Open Journal of Power Electronics*, vol. 6, pp. 883–898, 2025.

- [27] M. Cheng, W. Qin, X. Zhu, and Z. Wang, “Magnetic-Inductance: Concept, Definition, and Applications,” in *IEEE Transactions on Power Electronics*, vol. 37, no. 10, pp. 12406–12414, Oct. 2022.

- [28] B. Sauter, S. Reese, S. Sinha, J. Haddon, T. Byrd, and D. Maksimovic, “Integrating equation-based methods with random forest regression for improved accuracy of magnetic core loss modeling,” in *2024 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Long Beach, CA, USA, 2024, pp. 410–415.



Sheng Qu received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2025, where he is currently pursuing the M.S. degree. His current research interests include modeling, control and optimal design of isolated bidirectional DC-DC converters.



Shutong Sang received the B.S. degree in electrical engineering from Dalian Maritime University, Dalian, Liaoning, China, in 2024. He is currently working toward the M.S. degree in electrical engineering in Xi'an Jiaotong University, Xi'an, Shaanxi, China. His current research interest includes modeling of high-frequency magnetic materials.



Chuanyang Wang is currently working toward the B.S. degree in physics with Xi'an Jiaotong University, Xi'an, China. His research interests include artificial intelligence-assisted device loss modeling, automated device testing, and simulation.



Yuqi Wei was born in Henan, China, in 1995. He received his B.S. degree in electrical engineering from Yanshan University, Hebei, China, in 2016, and his M.S. degree in Electrical Engineering from University of Wisconsin-Milwaukee (UWM), Wisconsin, U.S.A., in 2018. He received another M.S. degree in electrical engineering from Chongqing University, Chongqing, China, in 2019. He was a visiting researcher with Kiel University, Germany from September to December 2021. He received his Ph.D. degree in electrical engineering from the University of Arkansas, Fayetteville, U.S.A, in 2022. He was a Postdoctoral researcher at University of Arkansas in 2022. Since 2023, he has been a full professor with Xi'an Jiaotong University. His current research interests include wide band gap devices, design automation and cryogenic power electronics.



Dongxing Yang received the B.S. degree in electrical Engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2025, where he is currently pursuing the M.S. degree. His research interests include high-frequency planar magnetics and high-density high-efficiency DC-DC converters.