

MagNet Challenge Final Submission Report

Emmanuel Havugimana Vivek Thomas Chacko Sritharini Radhakrishnan Jorge Peralta Mike Ranjram
ehavugim@asu.edu vtchacko@asu.edu sradha14@asu.edu japeral2@asu.edu Mike.Ranjram@asu.edu

Abstract—Participating teams of the 2023 MagNet Challenge compete to create a software package that utilizes the magnetic characteristics of different materials to predict their core loss efficiently and with high accuracy. The difficulty of this challenge is in using the developed model to accurately predict the core loss of new materials that it has not previously trained on. Over the course of the past few months, the ASUMag team has carried out exploratory analysis of the core loss computation equations and applied data-driven machine learning (ML) algorithms to develop a novel core loss-prediction model. Preliminary results were shared in our Nov. 1st submission. In our final submission, we refine our model to reduce its size and execution time significantly. Our main aim in this challenge is to produce a model of reasonable accuracy while minimizing its size. We believe we have successfully achieved this goal through the model design choices described below. In the final data set, we observed a subpar performance of our model on new material, noticeably a logarithmic relationship between error and data size. We attribute this observation to the size difference in the provided training sets from different materials than material differences. The data provided to train the model was significantly smaller than the data provided to the pre-test phase, which we assume to be the root cause of our model's performance, especially since any error in our model in training is only amplified in testing due to this data size imbalance. Our model's performance will revert to its nominally accurate core-loss predictions with a larger or more representative training data set.

Index Terms—core loss, MagNet challenge, Logarithm, re-sample, Feed Forward Neural Network

I. INTRODUCTION

In this paper, a model is proposed and compared with baseline models from MagNet AI [1] and equation-based models from MagNet-hosted tutorials [5]. Specifically, the ASUMag model is compared with accuracy limits stipulated by the 2023 MagNet Challenge [6]. The results are promising. Our models for each material share the same architecture, only differing in weight. Our material-independent architecture results from the inherent similarity of nominal core loss prediction, the physical basis of our loss-informed model architecture. Such modeling allows for cross-material learning as material extension becomes easier as one model is optimized. This allows for material similarity analysis and model versioning, which only involves updating the weights. These model weights are shared as JSON files for human readability and cross-platform support, making them easy to use. The model shared in the report is represented in a form that makes it easy to transfer across platforms.

II. LITERATURE REVIEW

The Princeton team reported errors based on the B-H loop and core loss [1]. The model results are compared with models

from Princeton team papers mainly [1]. The 95th percentile is considered equivalent to 2σ where standard deviation σ was reported instead of the 95th percentile and used half-normal distribution [4] approximations for deriving metrics from other metrics. This allows for the comparison of results from the literature and MagNet competition results. Even in cases where a subset of MagNet competition metrics are reported, a proxy can be computed for other metrics.

The results indicate a general model with performance that scales with training data. The estimates of the 95th percentile are from comparing the prediction of the model to the provided training set.

III. MODEL DESCRIPTION

A description of ASUMag's current model is given in this section. The model is based on 2023 MagNet-held tutorials. A unique contribution of introducing down-sampling in the data processing step allows for the reduction of model size while incorporating arbitrary waveshape modeling using the Feed Forward Neural network architecture. The limits of down-sampling was set by studying the reconstruction accuracy observed and our target model performance.

Our design makes our model more independent of the 'B' wave shape and allows it to be fitted with all 'B' wave shapes or selectively by one wave shape. As we have trained the model with all wave shapes, one model is shared for a material.

- 1) Input to model (29x1)
 - a) First 24 points sample of "B" (24x1)
 - b) Temperature (4 temperature) one-hot encoding (4x1)
 - c) Frequency (1x1)
- 2) Model: Feed Forward Neural Network with
 - a) Layer 1 : 29x29, Layer 2: 29x15, Layer 3: 15x15, Layer 4: 15x1
- 3) Output: 1x1 core loss

IV. MODEL SIZE AND PERFORMANCE ESTIMATES

The setup is a pipeline for reading data from the folders and evaluating a model with both RMS of absolute percent error and 95th percentile absolute percent error.

The 95th percentile error is computed for a model we developed for ten different materials.

This might be a result of material characteristics, data set size, waveform type, or composition ratios. Based on our exploration of model performance variation among the ten training materials since our Nov. 1st submission, we found

that it is redundant to develop different models for different materials.

We note that if there is any variation in error between the different materials for the same model, this might be a result of material characteristics, data set size, waveform type, or composition ratios. Table I shows our estimated performance per wave shape type and material. Based on our exploration of model performance variation among the ten training materials since our Nov. 1st submission, we found that it is redundant to develop different models for different materials.

Material	A	B	C	D	E
Model size	1576	1576	1576	1576	1576
Training data size	2432	7400	5357	580	2013
Estimate 95th perc. Err.	7.9%	5.3%	8.2%	19.3%	10.2%
Sine 95th perc. Err.	18%	8%	20%	26%	21%
Tri 95th perc. Err.	6%	5%	7%	17%	10%
Trap 95th perc. Err.	6%	5%	7%	28%	9%

TABLE I

MODEL SIZE FOR EACH MATERIAL AND ESTIMATES OF 95TH BY WAVE TYPE

V. ASUMAG MODEL CHOICES

A. Logarithmic transformations

The training loss of a model is usually reported as an absolute measure and is not relative to the size of the training data set. Therefore, we have chosen to apply a logarithmic transformation to the output of the model so that the resulting error is independent of the core loss magnitude.

This choice enables our model to meet the competition requirements of reporting percentage error without writing an additional custom loss function. Moreover, as Steinmetz's equation includes multiplicative and exponential parameters, a logarithmic transformation is suitable for model convergence besides the direct mapping of error in training and the percentage error. Although these transformations may limit model learning in the case of additive 'B' losses, we don't expect these effects to dominate in the test cases.

As a result of the logarithmic transformations we apply, we expect most of the errors during training to be less than one, which can be used to shape a histogram of errors by the custom error criterion. The observed distribution of error that seems to be size-dependent is likely due to possible data collection-related errors which inversely correlate to the measured size. As stated above, the log transform makes the percentage error size-independent; therefore, any residual 'dependence' still observed may result from the collected data's unpredictability. To rectify such anomalies requires a minor but more complex fitting trend.

From the logarithm transform,

$$value = \log(Power)$$

So, an error criterion (mean) during training becomes a percentage metric in the final prediction after anti-log

$$Power_{meas} = 10^{value}$$

$$Power_{pred} = 10^{value + criterionLoss}$$

$$Error = \left| \frac{Power_{pred}}{Power_{real}} - 1 \right|$$

$$Error = |10^{0+criterionLoss} - 1|$$

from $d \frac{d(10^x)}{dx} = \ln(10) * 10^x$, and from expansion of $10^{0+error}$ to first order term, we get

$$Error \approx |1 + \ln(10) \times criterionLoss - 1|$$

$$Error \approx \ln(10) \times criterionLoss$$

This direct link between percentage error and criterion loss can be exploited in training to shape the loss histogram. We think it can be done but have not implemented this in our model.

B. The 95th percentile error highly correlates with mean error

As reported in pre-test estimates, we found a consistent scaling factor of the 95th percentile error with respect to the mean error. Hence, during optimization, we did not create a new training loss for the 95th percentile metric. An error distribution can be skewed to be more 95th-percentile friendly, but not by much. Assuming a normal distribution of errors and a zero average mean (bias), the absolute error distribution gives a guide to transforming error reports missing the 95th percentile for comparison with the competition metric.

C. Simplified model structure

Our goal in the competition is to minimize the size of our model while still achieving reasonable results in accuracy. To achieve this goal, we have chosen to track only a few key parameters as dictated by the model equation per new material. We argue that one must also consider model storage when considering model size. Hence, one can save on model storage space by reducing the number of parameters to store when training a new material.

1) *Same data processor for all materials:* Material-dependent data transformations make data more palatable to train the model, but one must store normalization factors to remember the original form of the data, thus increasing the model size. Without such transformations, computation times do not necessarily get longer; however, we may sacrifice some model accuracy because we don't normalize the data.

All our model information is in the model weights; to add a new material, only the model weights must be added.

2) *Transfer learning:* With the same model structure, the material cross-learning is improved through initializing weights with other material models via material similarity analysis as in table III.

In table III, the 95th percentile is highly dependent on the trained material and variance across model initialization. The small error, when initialized with a material, might indicate some similarity between the two materials. This shows that with a growing repository of materials, new materials would easily be found by finding similar materials to initialize weights or do other training techniques with.

TABLE II

New Material /Old material	3C90	3C94	3.E6	3F4	77	78	N27	N30	N49	N87	Min	Max
A	13%	12%	11%	17%	8%	12%	13%	11%	12%	10%	8%	17%
B	6%	4%	7%	5%	7%	5%	5%	5%	5%	6%	4%	7%
C	9%	8%	8%	8%	7%	8%	9%	7%	9%	8%	7%	9%
D	27%	19%	21%	21%	22%	19%	23%	26%	32%	21%	19%	32%
E	18%	17%	14%	17%	14%	13%	15%	17%	17%	17%	13%	18%
Average	15%	12%	12%	14%	12%	11%	13%	13%	15%	12%	10%	17%

TABLE III

MODEL 95TH PERCENTILE GIVEN INITIALIZATION WITH EACH OF TEN MATERIALS

D. Model accuracy and data pre-processing error

1024 to 24 B_field transform: Our data processing stage involves down sampling 1024 data points to 24; this introduces a 2% 95th percentile error in the reconstruction of the 'B' wave shape. This was measured by reconstructing the original 1024 data points from the down-sampled 24 and comparing the differences as

$$\text{ReconstrError} = \frac{\sum(\text{abs}(\text{Breconstr} - \text{Boriginal}))}{\sum(\text{abs}(\text{Boriginal}))}$$

B^β , the error introduced, is approximately $2\% \times \beta$ approx 4% to 6%. Note: this can be reduced by using shifts of original data and re-sampling, sending all sampled shifts to the model, and looking at the results. This limitation can be traded with model size; we chose to reduce input size to reduce model size, which was acceptable given that the measurement accuracy are negligible. Our current model structure allows a designer to adapt it to be more accurate by increasing the size while keeping the architecture the same. The transformation is general in the sense it does not rely on concepts of duty cycle, zero voltage times, etc., which would assume a voltage-driven system (defined wave shape).

E. Treatment of temperature in the model

Currently, the model uses four inputs for temperature. For now, it uses one-hot encoding to encode 25, 40, 70, and 90 (the standard temperatures in the model). As a result, the model is limited by temperature in training. During testing, one can post-fit arbitrary temperatures by finding the losses at all temperatures and fitting the temperature afterward. For example, core loss at 30 degrees can be estimated by calculating loss at 25, 50, 70, and 90 degrees and then using interpolation to find the core loss at the specific temperature. We have not encountered any temperature values outside of what we have encoded during the training and testing. The one-hot temperature encoding was removed to incorporate arbitrary temperature training in the model. But given how temperature is weighed in the model as compared to 'B' or frequency, we think it might not be as useful to know core loss accuracy up to a 1-degree resolution. However, we believe the temperature encoding, as related to our model, can be improved in the future.

F. Model limitations and guidelines

This section states how not to use the model and describes potential pitfalls in the model to prevent misusage or misattribution of erroneous results.

Our model was trained on the MagNet dataset; hence, any limitations that the data might have in the loss's predictability will propagate to our model's results. The model structure should be trained and tested on the same data structures. We caution the user to use the same parameter units as the model as our current pre-processing stages do not have a unit detection system.

In the data set, the data provided was not phase-shifted by the competition organizers and thus mostly started at the same point or same phase for each shape type. We recommend keeping the wave shapes that way to preserve model accuracy. We noticed that the data wave shapes generally start with an increasing slope of 'B' and end with a decreasing slope of 'B'.

A pre-processor function has been added to shared scripts to ensure that is the case for the input 'B' field. Note that an attempt was made to make the model robust to shifts, and we think it would work if trained with shifted wave shapes, but to ensure consistency, we would prefer it if the user inputs data that adheres to the same format as the training data set. This pre-processor is informed by current knowledge that the representation of the 'B' wave shape shifts should not change the core loss.

G. Taking the geometric mean

We recommend taking a geometric mean when averaging results from the model for accuracy improvement purposes at a specific operating point from an averaging reducing variation perspective. We noticed improvements when using a geometric than an arithmetic mean, but have not done extensive studies on its usage to improve performance. The geometric mean is from the fact that the model is multiplicative in its effects from the logarithm transformation involved.

One potential application is using multiple versions of the model and averaging the results for each point. In that case, we recommend to consider taking a geometric mean. Averaging of the model results would improve the model standard deviations, thereby reducing the 95th percentile or long tails. Shared results don't use averaging of multiple versions as we would need to submit all models, thus increasing the size, but we think it would improve some benefits in an application setting dependent on how uncorrelated the errors are.

VI. MODEL DEPLOY-ABILITY AND USABILITY

Accompanying the report, the model is shared in JSON format with accompanying data transformations implemented in Python with the help of the NumPy package for matrix algebra, making it easy to transfer to other computing platforms. The model function as shared is more usable and easy to understand by anyone with an understanding of matrix algebra or any simulation software supporting matrix algebra.

To use the model in simulation software, one would need a 24-point sampler of the 'B' wave shape. The sampled wave shape concatenated with the temperature and frequency parameters would be fed to a series of matrix multiplications (weights and bias) and threshold (IF statement) or gating for ReLU implementation; changing material would amount to changing the matrices. Note: we have not provided a sample implementation.

VII. CONCLUSION

The ASUMag team has developed a model that has a comparable performance to the model proposed in [1] with a lower model parameter/model size. The results from this model are used across materials with varying data sizes and compositions.

The model results and model function accompany the report as per competition requirements. Overall, our model structure is capable of arbitrary wave shape prediction. The model is explained, and its application is addressed.

REFERENCES

- [1] Chen, Minjie; Li, Haoran; Serrano, Diego; Wang, Shukai (2023): MagNet-AI: Neural Network as Datasheet for Magnetics Modeling and Material Recommendation. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.22726115.v1>
- [2] D. Serrano et al., "Neural Network as Datasheet: Modeling B-H Loops of Power Magnetics with Sequence-to-Sequence LSTM Encoder-Decoder Architecture," 2022 IEEE 23rd Workshop on Control and Modeling for Power Electronics (COMPEL), Tel Aviv, Israel, 2022, pp. 1-8, doi: 10.1109/COMPEL53829.2022.9829998.
- [3] Chen, Minjie (2022). How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.21340998.v3>
- [4] "Half-normal distribution," Wikipedia, https://en.m.wikipedia.org/wiki/Half-normal_distribution (accessed May 31, 2023).
- [5] Otvam, "Otvam/magnet_webinar_eqn_models: 2023 Magnet Challenge Webinar: Equation-based Baseline Models," GitHub, https://github.com/otvam/magnet_webinar_eqn_models(accessed May 31, 2023).
- [6] Minjiechen, "Minjiechen/magnetchallenge: Magnet challenge," GitHub, <https://github.com/minjiechen/magnetchallenge> (accessed May 31, 2023).

CU Boulder MagNet Challenge Final Report

Bailey Sauter*, Skye Reese*, Shivangi Sinha*,

*Ph.D. Students

Colorado Power Electronics Center

University of Colorado Boulder

Email: {bas4247, skre2597, shsi3480}@colorado.edu

Abstract—This report summarizes the work of the University of Colorado at Boulder (CU-Boulder) team on the IEEE PELS MagNet challenge to predict core loss in magnetic components. This report presents a novel approach that integrates an existing equation-based core loss algorithm based on Steinmetz parameters with a simple random forest regression model to provide greater accuracy in core loss predictions without incurring significant computational costs. This hybrid model uses the equation-based model as a starting point and attempts to predict and correct the error through a multiplicative correction coefficient. The regression that produces the correction coefficient is trained using the MagNet database of experimentally measured core loss data, which includes a wide variety of materials and operating conditions [1]. Results for the ten original MagNet challenge materials are provided alongside test results on the five new, unidentified materials (labeled as A-E). Results across all materials showed an average 4x reduction in percent error estimating volumetric core loss.

I. INTRODUCTION

The MagNet [1] database offers a large amount of experimentally measured core loss samples available for validation and training. This enables a more in-depth look into alternative data-driven core loss modeling approaches that expand on existing, equation-based models like GSE [2] and iGSE [3]. Recent work has explored the use of neural networks, of both deep and convolutional types, for this application [4]–[6]. However, neural networks can be quite complex to set up and train and can require significant computational resources along with large input vectors.

This report introduces a method that improves upon the previously developed equation-based models by adding a relatively simple machine learning (ML) based regression with easy-to-obtain input features and using it in a hybrid model to predict a multiplicative correction factor for the iGSE algorithm. Specifically, a random forest regression is chosen for this task due to its simplicity, accuracy, and resistance to overfitting [7]. The result is a simple, computationally inexpensive, hybrid model that provides significant improvements

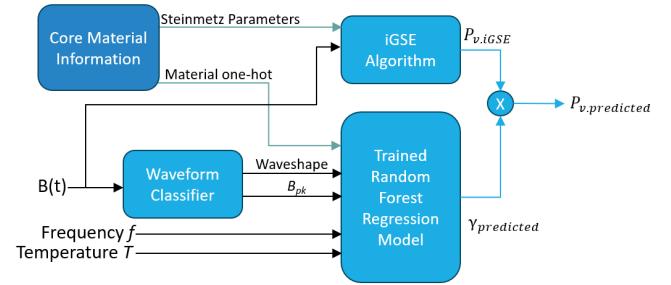


Figure 1: A flowchart illustrating the structure of the CU-Boulder hybrid model: core loss predictions obtained using the iGSE algorithm are corrected via a multiplicative correction factor $\gamma_{predicted}$ generated by a regression trained on the core loss data samples in the MagNet database.

over standard equation-based approaches. A flowchart of the proposed hybrid model is shown in Fig. 1.

The report is organized as follows. Data pre-processing aimed at waveform classification and optimization of Steinmetz parameters for sinusoidal waveforms is addressed in Section II. Section III summarizes the results of core loss prediction on the database using existing, equation-based methods. Section IV presents the results of the novel hybrid model, its improvement over existing standard algorithms, and its potential for use in the power electronics design process. Section V summarizes the results and concludes the report.

II. DATA PRE-PROCESSING

The database used to train and evaluate the model in this report is the database of core loss samples provided by the research group at Princeton University [1]. The initial data is made up of 186,757 core loss samples of ten ferrite materials. The materials are all commonly available and range in designed frequency from below 100 kHz to an upper bound of 2 MHz. The tested operating frequencies within the database range from below 50 kHz to greater than 600 kHz. The sampled temperatures range from 25°C to 100°C. The database includes sinusoidal, triangular, and trapezoidal flux-density waveforms, of varying slopes and duty cycles.

After the initial model was designed and trained using the aforementioned dataset, limited data for five additional materials, labeled A-E and of unknown identity and type,

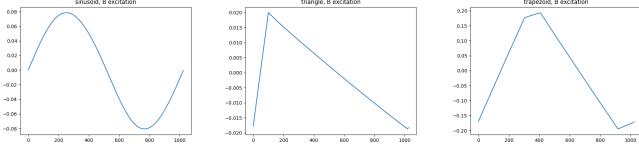


Figure 2: An example of sinusoidal, triangular, and trapezoidal waveform types respectively, all classified correctly by the algorithm.

was provided and incorporated into the model’s training. This limited data included 17,782 new samples that could be used in training, with an additional group of samples (core loss not provided) also present for final scoring of the model performance within the MagNet challenge.

A. Waveform Classification

In order to gain more information about the core loss samples in the database, an algorithm was implemented to classify the samples in the database based on the waveshape of their flux-density waveforms. All the flux-density waveforms in the database of samples [1] fall into one of three categories: sinusoidal, trapezoidal, or triangular waveshapes, which are the most common flux-density waveforms in switched-mode power converters. Therefore, these are the categories used and referenced in the rest of the report.

The classification of sinusoidal waveforms is performed by first taking the Fourier transform and analyzing its power spectral density. A power spectral density value above .499 indicates a sinusoidal waveform. For non-sinusoidal waveshapes, the derivative of the waveform is taken. The waveform is split up into segments based on when the slope changes. The number of segments is analyzed in order to differentiate between triangular and trapezoidal waveforms (2 segments for a triangular waveform, 4 segments for a trapezoidal waveform). Examples of each category of waveshape are shown in Fig. 2.

A flux-density waveform classification accuracy of >99% was measured on the limited dataset (materials A-E). Waveshape information was not provided on the initial, larger dataset, but accuracy can be assumed to be similar based on visual inspection of random samples and given that the waveshapes should not depend on the material.

B. Optimization of Steinmetz Parameters for Sinusoidal Data

The Generalized Steinmetz Equation (GSE), Eq. (1), attempts to predict volumetric core loss P_v in [W/m^3] (with sinusoidal excitation) using a set of simple parameters for a single core material regardless of temperature, excitation, and frequency [2].

$$P_v = k \cdot f^\alpha \cdot B_{pk}^\beta \quad (1)$$

The iGSE core loss algorithm [3] improves on Eq. (1) by breaking up waveforms into smaller subsections but uses the same input parameters. In reality, Steinmetz parameters vary based on the operating conditions and the accuracy of datasheet measurements. This introduces significant errors when selecting the Steinmetz parameters from arbitrary operating conditions in the datasheet. Given the high volume of

Model Size	774 MB (.joblib file)
Number of Parameters	100 estimators (hyperparameters incl. in text)

Table I: Regression model information

core loss samples available in the database [1], the optimal Steinmetz parameters that minimize the error in the GSE core loss calculation are found numerically, giving the ML model a starting point with a higher accuracy than hand-calculated Steinmetz parameters from a datasheet. Only samples with sinusoidal flux-density waveforms are used in this optimization because the GSE technique is explicitly designed for only sinusoidal excitation waveforms. To optimize the Steinmetz parameters, the log of Eq. 1 is taken and an Ordinary Least Squares (OLS) model is used to solve the multiple linear regression for optimal values of k , α , and β for each material in the database. The solver is limited by the constraints $1.6 \leq \alpha \leq 2.0$ and $2.3 \leq \beta \leq 2.7$, a reasonable range of standard Steinmetz parameters, so that the GSE fits are similar across the materials. These parameters are used as an input to the equation-based model [3] in the following sections.

III. RESULTS FOR EQUATION-BASED MODELS

After the classification of samples by waveshape and the optimization of Steinmetz parameters, initial results were generated by the GSE [2] and iGSE [3] core loss models. The mean absolute error (%) for both approaches can be seen across various materials in Table II. As expected, Table II shows that iGSE offers more accurate results than GSE and is therefore used as the starting point for the machine-learning (ML) based correction. To better understand the dataset and under what conditions the errors in existing core loss models are largest in magnitude, the error data seen in Table II for the iGSE algorithm is displayed on a heatmap in Fig. 4 plotted against the frequency and excitation magnitude with the number of points at that operating condition shown via the size of the plotted circle. The heatmaps included in the report are only based on the initial training data that is complete (10 known materials). It should be noted that in order for the heatmap to be readable, similar frequency and flux-density values are grouped together and have their absolute percent error averaged and displayed at one point.

IV. HYBRID MODEL AND RESULTS

The structure of the hybrid model proposed in this report is shown in Fig. 1. Errors in core loss predictions obtained using the iGSE algorithm are corrected via a multiplicative correction factor $\gamma_{predicted}$ generated by a regression trained on the core loss samples in the MagNet database [1]. The use of one-hot encoding of the material enables the use of a single model for all materials. This optimizes usability and space while improving model versatility to allow for good performance on new materials not included in the training dataset [8].

A. Hybrid Model Structure

The regression model used in the machine learning portion of the hybrid model shown in Fig. 1 is a random forest regression model implemented using the scikit learn ensemble package in Python [9]. Random forest regression models are made up of an ensemble of decision trees [10]. Decision trees make predictions using a series of splits in the training data that are based on a sample's value of a given feature. Each decision tree is trained on a random subset (with replacement) of the training data. The resulting prediction of the random forest model is an average of the predictions of the trees that make up the ensemble. The decision tree structure of each estimator inherently captures relationships between variables and gives outlier points less influence over the model. Meanwhile, the ensemble nature of the model reduces its susceptibility to overfitting during training. These advantages make a random forest regression particularly suitable for the task of core loss estimation, which has been shown to be highly nonlinear, and not easily predictable via physics-based explanations.

The machine learning portion of the hybrid model uses the results from Section III as a starting point. A correction factor, γ , where

$$\gamma = P_{v.\text{true}} / P_{v.iGSE} \quad (2)$$

is calculated for each sample. This is used as the target output variable y in the training of the random forest regression model. The model was trained on 80% of the available samples and tested on the remaining 20% (of 204,539 total core loss samples across the fifteen materials). Within the training set, 10,000 samples were taken (with replacement) from each of the fifteen material categories. This method prevents overfitting to materials with more samples taken and improves performance on the new, unknown materials. The features used as inputs for the model are one-hot encoding of the material, the Steinmetz parameters of that material, temperature, peak flux density, frequency, and a one-hot encoding of the waveshapes as classified in Section II. The peak flux density is defined as

$$B_{pk} = \frac{\max(B) - \min(B)}{2} \quad (3)$$

where B is the flux density waveform of the given data sample. Features were normalized and standardized to mitigate magnitude variation before training.

After the model is trained, it is used to predict $\gamma_{predicted}$ for the test set, which is then used to calculate the predicted P_v ,

$$P_{v.predicted} = \gamma_{predicted} \cdot P_{v.iGSE} \quad (4)$$

as shown in Fig. 1. The absolute percent errors are then calculated and compared with the results of Section III.

Table I shows the size of the model (used for all fifteen materials) after being saved as a .joblib file for future use. In the random forest regression, the hyperparameters used are $n_{estimators} = 100$ (the number of trees in the ensemble) and $depth_{max} = 30$ (maximum depth of each tree). Limiting the maximum depth of the trees helps prevent overfitting of the

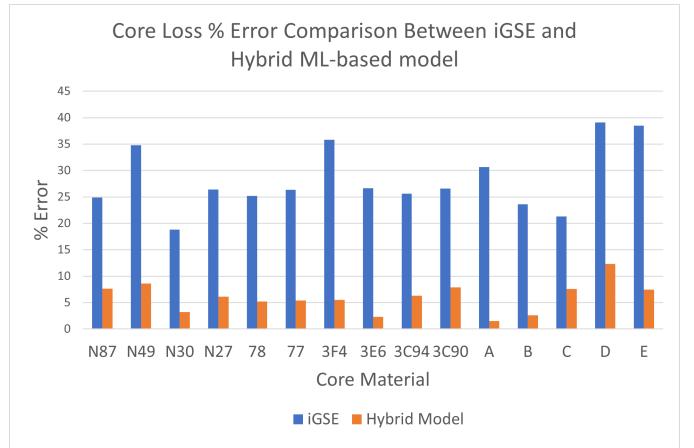


Figure 3: Histogram comparing the mean core loss % error across the 15 tested materials using the iGSE method and using the hybrid model.

model to the training data. A higher number of estimators has the potential to increase the accuracy of the model on more complex data, but, in this case, performance was found to plateau at $n_{estimators} = 100$. Mean squared error was used as the split criterion during the training of the decision trees.

B. Hybrid Model Results on Test Dataset

The results of the hybrid model are presented in Tables II and III (average and 95th percentile, respectively). They show a marked improvement in accuracy compared to the equation-based models. Across all materials, the mean absolute percent error is 4x less than what it was using only the iGSE method.

A histogram comparing the results of the core loss prediction algorithm to the conventional iGSE algorithm across the ten different materials present in the dataset is shown in Fig. 3. In addition, Fig. 5 is a heatmap showing the percent error in core loss predictions generated by the hybrid model across all materials. This offers a direct comparison to Fig. 4. It can be seen that the hybrid model reduces the overall error in core loss prediction minimizes the concentration of errors at certain operating conditions. This indicates that the model is more reliably accurate across the wide variety of operating conditions present in the dataset.

Table IV is a table of the hybrid model results separated by waveshape. In all cases, across material and waveshape, the hybrid model reduces error from the iGSE model. Most flux density waveforms in power electronic circuits are non-sinusoidal. Therefore, this table demonstrates the usefulness of the proposed model's versatility across operating conditions.

In summary, the model performance on the test dataset validates the use of the hybrid model to predict core losses in switched-mode power converters. Large improvements are seen across a wide variety of materials, operating conditions, and waveshapes, demonstrating that the model is not limited to specific use cases but has the potential to serve as an effective general model for core loss. The model's performance on new ferrite materials not included in the training set demonstrates its versatility and is reported in a paper to be presented at IEEE APEC 2024 [8].

Table IV: Mean Absolute Percent Error of Models by Material, Separated by Waveshape Type

Table II: Mean Absolute Percent Error of Models by Material

Material	GSE Error [%]	iGSE Error [%]	Hybrid Model Error [%]
N87	29.13	24.93	7.62
N49	22.49	34.81	8.63
N30	26.33	18.81	3.20
N27	30.32	26.46	6.10
78	30.98	25.20	5.28
77	31.85	26.35	5.41
3F4	17.24	35.84	5.56
3E6	25.52	26.63	2.38
3C94	30.82	25.58	6.36
3C90	31.32	26.62	7.90
A	36.65	30.68	11.51
B	29.68	23.63	2.66
C	30.25	21.28	7.62
D	39.38	39.09	12.33
E	40.38	38.46	7.46

Table III: 95th percentile Absolute % Error of Models by Material

Material	GSE 95th Error [%]	iGSE 95th Error [%]	Hybrid Model 95th Error [%]
N87	70.21	57.06	27.80
N49	91.91	89.78	31.67
N30	72.46	48.49	9.71
N27	72.11	59.81	21.96
78	73.84	60.18	17.10
77	74.31	62.55	18.51
3F4	89.31	86.61	18.10
3E6	79.71	81.24	7.39
3C94	74.08	57.98	21.88
3C90	74.92	62.44	28.77
A	81.92	64.30	38.61
B	70.31	55.47	7.69
C	77.77	58.03	25.79
D	140.20	137.49	35.62
E	121.98	145.31	24.57

Material	Waveshape	iGSE Error [%]	Hybrid Model Error [%]
N87	Sinusoid	30.06	5.10
	Triangle	26.43	9.51
	Trapezoid	23.84	6.73
N49	Sinusoid	41.38	12.85
	Triangle	35.96	9.81
	Trapezoid	33.72	7.67
N30	Sinusoid	18.75	4.53
	Triangle	19.96	2.49
	Trapezoid	18.26	3.43
N27	Sinusoid	28.33	4.50
	Triangle	27.07	6.73
	Trapezoid	24.45	5.88
78	Sinusoid	27.46	3.32
	Triangle	26.90	5.68
	Trapezoid	23.65	5.18
77	Sinusoid	30.12	5.03
	Triangle	27.37	6.23
	Trapezoid	25.55	5.00
3F4	Sinusoid	37.29	4.82
	Triangle	39.26	6.32
	Trapezoid	33.80	5.16
3E6	Sinusoid	22.56	2.52
	Triangle	29.56	2.23
	Trapezoid	25.68	2.44
3C94	Sinusoid	28.30	4.62
	Triangle	26.98	7.14
	Trapezoid	24.61	6.06
3C90	Sinusoid	29.93	6.75
	Triangle	27.88	9.22
	Trapezoid	25.74	7.22
Material A	Sinusoid	39.16	26.86
	Triangle	33.82	13.49
	Trapezoid	28.72	9.69
Material B	Sinusoid	23.43	4.41
	Triangle	23.79	2.68
	Trapezoid	23.57	2.52
Material C	Sinusoid	31.84	17.16
	Triangle	22.81	8.15
	Trapezoid	19.75	6.67
Material D	Sinusoid	36.11	10.89
	Triangle	40.96	12.59
	Trapezoid	21.07	14.79
Material E	Sinusoid	18.26	9.79
	Triangle	44.72	8.38
	Trapezoid	35.66	6.74

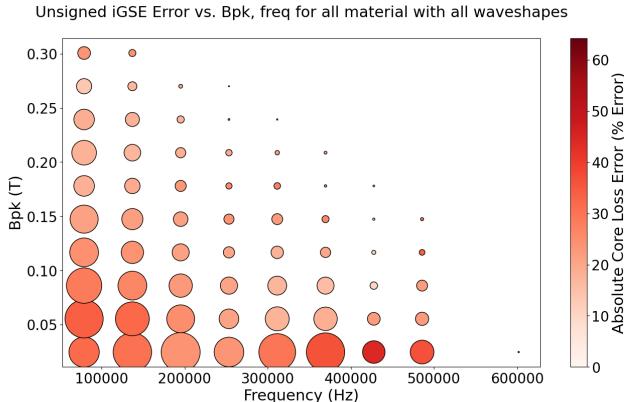


Figure 4: An error heatmap of the iGSE core loss model shows percent error on a color scale presented on an axis of peak flux density and frequency. The size of the point indicates the ratio of samples at that operating condition.

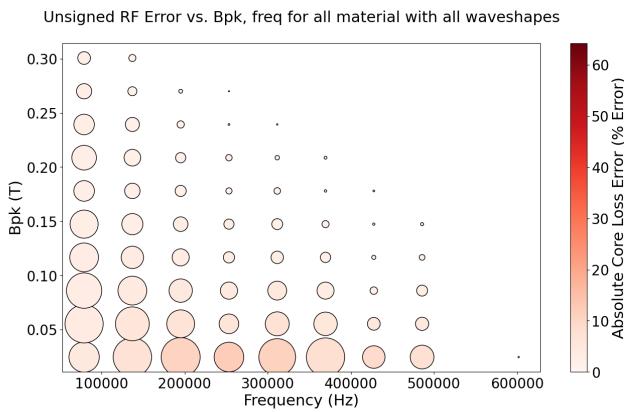


Figure 5: An error heatmap of the hybrid core loss model shows percent error on a color scale presented on an axis of peak flux density and frequency. The size of the point indicates the ratio of samples at that operating condition. The color scale shown is the same as in Figure 4 to offer a direct comparison.

C. Limitations and Future Work

The flux-density waveform for all the data used to train the hybrid model presented in this paper has a DC bias of zero. Therefore, future areas of exploration may include the incorporation of DC bias into the model (shown to alter core loss in [11], [12]). Further improvements in accuracy can also be gained by including permeability as a feature if it is known for each material.

In practice, the proposed hybrid core-loss model can be used independently, simply to improve core-loss predictions in the context of the design of magnetic components. Other applications of this model include the incorporation into a larger data-driven design tool such as that described in [13].

V. CONCLUSIONS

Equation-based approximations for modeling of core loss, such as [2], [3] have been widely used in power electronics practice. However, due to the wide range of desired operating conditions and irregularities in core materials, a more accurate model is necessary to improve core-loss modeling and truly enable optimization of magnetic components in switched-mode power converters. Taking advantage of the large database

of experimentally measured core loss samples in [1], this report presents a novel hybrid model that integrates the iGSE method [3] with a waveform classifier and a relatively simple machine-learning based, random forest regression to improve the modeling of core losses across operating conditions. The hybrid model is shown to reduce absolute percent error by a factor of four across all materials using only simple and easily available features. This approach eliminates the necessity for overly complex and computationally expensive methods such as neural networks while still offering greatly improved and repeatable results.

The model has significant potential to serve as an effective general model for core loss. Its performance on new materials will be shown in an upcoming paper at IEEE APEC 2024 [8].

REFERENCES

- [1] H. Li, D. Serrano, T. Guillod, E. Dogariu, A. Nadler, S. Wang, M. Luo, V. Bansal, Y. Chen, C. R. Sullivan, and M. Chen, "MagNet: An Open-Source Database for Data-Driven Magnetic Core Loss Modeling," in *2022 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2022, pp. 588–595.
- [2] C. Steinmetz, "On the law of hysteresis," *Proceedings of the IEEE*, vol. 72, no. 2, pp. 197–221, 1984.
- [3] K. Venkatachalam, C. 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.*, 2002, pp. 36–41.
- [4] 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)*, 2022, pp. 1–8.
- [5] E. Dogariu, H. Li, D. Serrano López, S. Wang, M. Luo, and M. Chen, "Transfer Learning Methods for Magnetic Core Loss Modeling," in *2021 IEEE 22nd Workshop on Control and Modelling of Power Electronics (COMPEL)*, 2021, pp. 1–6.
- [6] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen, and M. Chen, "MagNet: A Machine Learning Framework for Magnetic Core Loss Modeling," in *2020 IEEE 21st Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2020, pp. 1–8.
- [7] J. K. Jaiswal and R. Samikannu, "Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression," in *2017 World Congress on Computing and Communication Technologies (WCCCT)*, 2017, pp. 65–68.
- [8] 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)*, 2024.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [10] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [11] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Core Losses Under the DC Bias Condition Based on Steinmetz Parameters," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 953–963, 2012.
- [12] B. N. Sanusi, M. Zambach, C. Frandsen, M. Beleggia, A. Michael Jørgensen, and Z. Ouyang, "Investigation and Modeling of DC Bias Impact on Core Losses at High Frequency," *IEEE Transactions on Power Electronics*, vol. 38, no. 6, pp. 7444–7458, 2023.
- [13] S. Reese, B. Sauter, S. Khandelwal, A. Kamalapur, T. Byrd, J. Haddon, and D. Maksimovic, "Loss estimation and design of dc-dc converters using physics- and data-based component models," in *2023 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2023, pp. 82–89.

MagNet Challenge 2023 - University of Bristol

Lizhong Zhang, Tom McKeague, Binyu Cui, Navid Rasekh,

Jun Wang (Academic Advisor), Song Liu (Academic Advisor), Alfonso Martinez (Industry Advisor)

{ts20212, rd20871, bc16461, ck19302, Jun.Wang, Song.Liu}@bristol.ac.uk; alfonso.martinez@we-online.com

Abstract—This report describes the work done by the University of Bristol Team in response to the MagNet Challenge 2023.

Keywords—*MagNet Challenge, machine learning, neural networks*

I. INTRODUCTION

Nowadays, magnetic components are involved in most power electronics applications for functionality and filtering purposes. It's typically known to be the least efficient component that has a massive impact on system performance and efficiency regarding the size, weight and power loss factor [1][2]. However, an accurate core loss model for magnetic components that is based on the first principle remains elusive due to the non-linear feature of the magnetic material and other intercoupled factors such as DC-bias condition. Numerous research studies have been carried out to factor in the external parameters contributing to magnetic loss under nonsinusoidal excitations. The Steinmetz equation (SE), shown in (1), is widely accepted as an empirical model to calculate core loss under sinusoidal excitation. The k, α and β are the SE parameters which can be calculated by substituting the measured core loss value with the corresponding frequency and peak flux density values. This results in diminished precision of the equation due to the SE parameters demonstrating inconsistent performance across different frequency ranges.

$$P_{loss} = kf^\alpha B^\beta \quad (1)$$

To enhance the universality and accuracy of the SE, the improved generalised Steinmetz equation (iGSE), as a modified solution, has been proposed based on SE for calculating core loss for arbitrary flux waveform under zero DC-biased condition [3]. The core concept of iGSE is to divide the complex waveform into individual B-H loops and calculate the loss respectively. However, these models generally face limitations in accuracy, particularly with certain waveform types, and tend to overlook the effects of temperature. Another approach called ‘Loss map’ is proposed in [4][5] for incorporating the pre-magnetization effects. To begin with, the operating state of one magnetic component can be described by three variables, the pre-magnetization state, the magnetic flux density swing and the flux density change rate. By measuring the B-H loop at various preset operating points, a core loss profile can be produced to cover all the operating conditions and used directly as a look-up table in further in-situ experiments.

In recent years, methods using neural networks and other machine learning techniques have demonstrated excellent results in addressing nonlinear regression problems and

forecasting time series data (e.g., image & voice identification) [6]. This technique could also be applied in core loss predicting and has proven to be more precise than the classic modelling based on the Steinmetz equation [7][8].

Three neural network models have been proposed and discussed in [7]. There is the ‘scalar-to-scalar’ model such as FNN which acts similarly to the Steinmetz equation that uses parameters such as flux density and frequency to directly predict the power loss. Unlike the limitations of SE, the FNN-based model is proven to have a higher accuracy rate across the frequency range while covering external influencing factors such as temperature. One drawback of this model is that different models have to be trained according to the excitation waveform type which the scalar could not fully represent. On the other hand, instead of using magnetic scalars, the ‘sequence-to-sequence’ model such as the transformer model [9] introduces the complete excitation waveform to the training process and predicts the magnetic response. While solving the issue of producing corresponding models to waveform types, the ‘sequence-to-sequence’ model normally introduces an enormous amount of parameters in both the input and output sides of the model which would lead to a longer training process and higher requirement for the training platform.

To leverage the strengths of the aforementioned models, the Long Short-Term Memory (LSTM) network which is a ‘sequence-to-scalars’ model has been chosen as our model for participation in the MagNet challenge. The LSTM network excels in processing regression problems with sequential input due to its key feature of capturing long-term dependencies in data and overcoming the short-memory issue prevalent in standard RNN models [10]. These characteristics are ideally suited to fulfil the requirements for processing and analysing the time series data relationships, such as those between $B(t)$ and $H(t)$.

In response to the MagNet Challenge 2023, this report details the approaches applied by the University of Bristol team for the competition. Section II gives an overview of the challenge. Section III introduces the machine learning framework and detailed considerations. Section IV shows the results and a summary of our models.

II. THE GOAL AND CHALLENGES

The aim of this project is to yield a “prediction model” for one magnetic core material, which takes in three inputs, B , f and T and outputs one volumetric loss density value. To achieve this aim, a machine learning process is expected to learn from the large database provided, which is experimentally measured loss and waveforms. This database is treated as the “ground truth” in

this work. Hence the top-level idea of this work is illustrated in Fig. 1

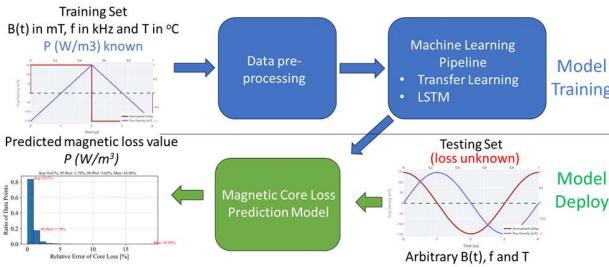


Fig. 1. The goal

There are a few challenges embedded in the training and testing datasets provided to us, which reflects real-life user scenarios. As illustrated in Fig. 2, there is a large-scale database of ten magnetic materials provided as the solid ground and starting point of this work. This large-scale database can serve as the foundation for the pre-training in a transfer learning process. Reflecting the real-world scenario to adapt a trained base model to a new material, the data available for five unknown materials are given with a small and unbalanced training set. The final testing data is given with inputs only to test out the model predictions.

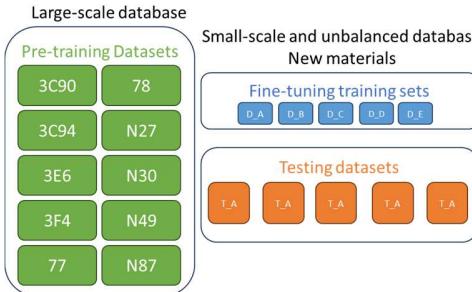


Fig. 2. The challenges in the given datasets

III. MACHINE LEARNING FRAMEWORK

A. Overall structure/pipeline with Transfer Learning

Given the rich training set for the ten known materials and limited data for the five unknown materials, this work has adapted a transfer learning pipeline, which contains a pre-training stage and a fine-tuning stage. This pipeline is illustrated in Fig. 3. Starting from zero in the pre-training stage, the machine learning model is trained by a large dataset of one particular material. As the most commonly-used and perhaps the most representative material in engineering, the training data of material 3C90 is used to generate a “fundamental model” as a generalised model to extract the key physical patterns for the general task of predicting the core loss value given the input data. The fundamental model is then tuned in the second-stage of pre-training against the ten materials’ training set, which yields 10 “base models” $\{M_i, i = 1 \dots 10\}$.

In the fine-tuning stage, one of the base model is selected and tuned specifically for one particular unknown material (e.g. Material A) in the final test data. This stage starts with a selection logic to pick out the best tuned base models based on the minimal average error that one base model yields on the testing data, which can be considered as identifying one material out of the original ten materials that is most similar to Material

A. At the end, a fine-tuned model for Material A is generated for deployment.

This pipeline is designed to make the most out of the rich training set of the original ten materials and transfer the learned knowledge into the fine-tuning step, which is a solution to the limited dataset for the testing data/materials. This idea is inspired from a machine learning concept called “meta learning”. The “meta” parameters of a neural network are pretrained using a similar material (3C90) and adapted to the target materials with further finetuning.

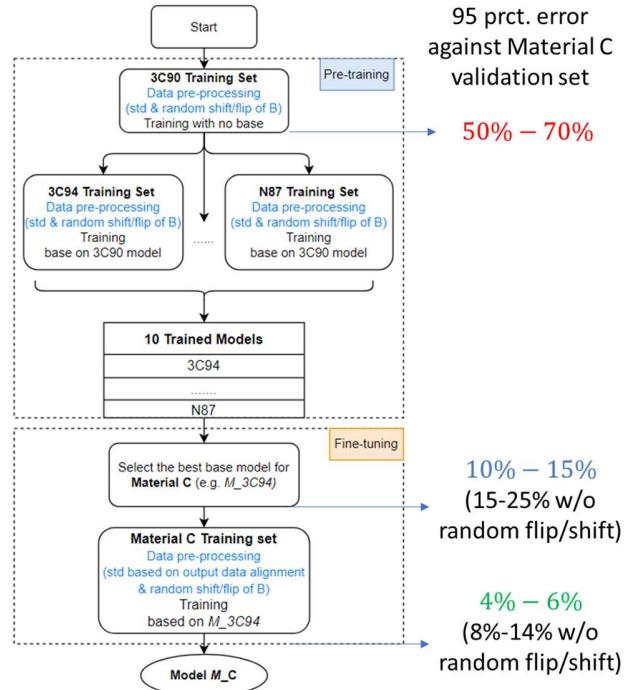


Fig. 3. Overall pipeline with transfer learning (example with 3C90 as the fundamental base and Material C as the target)

For the training process, the data are split into training set, validation set and a test set on a 70-20-10 basis. The split is performed randomly to minimize bias.

B. Data pre-processing

As marked in blue texts in Fig. 3, the training data has gone through a data pre-processing process before each training task takes place.

Data standardization

The input variables temperature and frequency are linearly rescaled into values in the range of [0, 1]. The B waveform and the loss density are also standardized so values are in the range of [-1, 1]. This procedure is to retain the polarity information and the linear correlation between the primary input (B waveform) and the single model output (loss density) for better LSTM waveform feature extraction. We observe that this data standardization process can enhance the model performance, avoids numerical instabilities and improves model performance. For the LSTM, data standardization can improve the convergence in the gradient descent optimization process. In the case of standardising the B waveform, it can be considered as extracting the shape of the waveform and minimize the numerical impacts of the magnitude of data. The standardization process is illustrated in Fig. 4, where the raw data is processed

into standardized data and stored together with the linear standardization coefficients (k , b) for each case. The scaling/standardization coefficients are determined for each material based the range of data, e.g. max value.

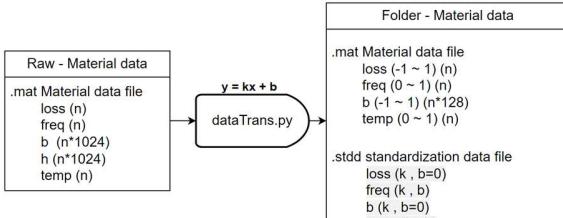


Fig. 4. Data standardization process (linear rescaling through $y = kx + b$)

Random shift/flip of B waveform

To enhance the generalisation and avoid overfitting to the time-sequence waveform data, a “random shift/flip” operation is implemented on the B waveforms before they enter the training stage. The assumption is that shifting horizontally (i.e. applying a phase shift) and/or flipping (vertically and horizontally) the B waveform will not change the core loss density associated with this waveform – based on domain knowledge, this assumption should be valid given the core loss density is an averaged value over one whole cycle of a periodic B waveform repeating itself once every 1/f second. This technique is similar to duplicating training images by rotating them in image recognition tasks. By injecting artificially rotated images, the neural network is trained to capture rotation invariant information from images. However, in our case, a rotation (pivoting) of the B waveform, instead of shifting or flipping, will invalidate the equal loss assumption. Note the phase shifting is similar to the approach in [11], while the flipping is an original contribution of this work.

To implement this operation, a function is applied on B waveforms to shift and/or flip them randomly before feeding them into the LSTM process as the sequential data input. Based on observations, this technique can significantly improve the model accuracy and reduce overfitting to the sequence of the signal.

Downsampling

The original 1024-point B waveform is downsampled to 128-point data based on a trial-and-error approach to reduce the computation need in the training process and model size while it still maintains a level of accuracy without missing key patterns/details. This operation may also reduce the overfitting on irrelevant/insignificant details contained in the waveforms.

Output data alignment for data standardization in the fine-tuning stage

To cope with the small-training-set problem, a special “output data alignment” technique is applied in the fine-tuning stage as a data standardisation approach. Given the objective in the fine tuning stage is to minimise the modification of the base model and unwanted overfitting towards the small training set, the output data (core loss density) to be fed into the training process is standardized in a different way – they are rescaled in a manner to align their extracted features with those of the original training data, to enhance the fine-tuned model. In order to correct the output shift, we apply shift and rescaling to y to match the distributions in the source/training set.

C. LSTM Neural Network Architecture

An LSTM + feedforward neural network (FNN) framework is applied as the fundamental machine learning approach. As shown in Fig. 5, LSTM is applied to deal with the B waveform only, excluding the temperature and frequency. This design enhances the feature extraction and pattern recognition of the waveforms to be better transferred to other models in the transfer learning architecture. The neural network generates one scalar, the volumetric magnetic loss, as the output – the whole process is a sequence-to-scalar model.

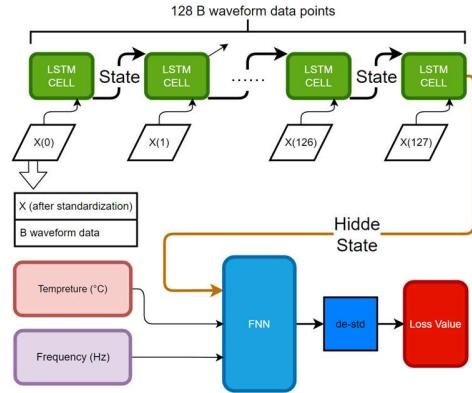


Fig. 5. Example of LSTM neural network structure

The loss function for the optimization process of the neural network is defined to minimize the relative error.

D. Data post-processing

Data de-standardization

Following the data standardization, we apply a reverse process to translate back the loss density value to the original scale at the end of the prediction pipeline.

Averaging output with random shift

When we deploy the model, shifting/flipping the B waveform can still lead to variations of the predicted loss. To take this factor into account, for each entry of the test data, the B waveform is processed into 100 (adjustable) instances with different phase shift/flips, which yields 100 predicted volumetric loss values. These values are averaged in a post-processing step to obtain the final prediction for this entry.

E. Data imputation

To cope with the cases with limited and unbalanced training set, we applied a data imputation approach to artificially expand the training set in these cases. For example, the training set of Material D only contains 580 entries, with an unbalanced split between sinusoidal, triangular and trapezoidal waveforms (e.g. 145/400/35). However, the testing data of Material D is 7299 entries with a 61/2247/4991 split – the testing data has a large portion with trapezoidal waves, while the training set has very limited data for this case. To compensate this mismatch, the trapezoidal data in the training set is artificially duplicated and expanded to match the waveform split in the test set as much as possible, which leads to a 145/2000/700 split.

IV. RESULTS AND SUMMARY

Based on the above machine learning framework, five models are produced for the five unknown materials. The performance of these models are evaluated against the validation

sets in the five cases, with the results shown in Fig. 6. Although the average of 95-percentile-error for the 10 known materials can achieve around 2.5%, the model's performance deteriorated in the cases of the 5 unknown materials due to the limited training sets. Material D is particularly problematic due to only 580 entries of training data available, which leads to a 22.04% 95-percentile-error. The average 95-percentile-error of the other four unknown materials achieves a 6.6%.

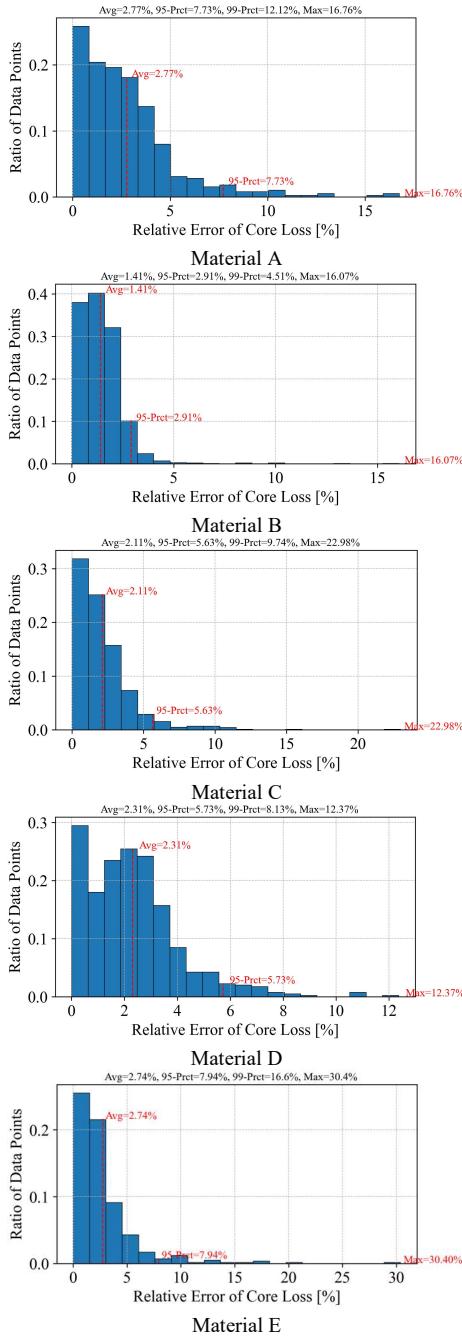


Fig. 6. Error distribution against the validation set for the target five unknown materials

The size of the five models are listed below in terms of number of parameters and file sizes. They all have the same size since our model architecture is fixed when training the base mode.

Table I Model Size

	Material A	Material B	Material C	Material D	Material E
Number of parameters	15,653	90,653	90,653	16,449	16,449
Model Size	361 KB	361 KB	361 KB	70 KB	70 KB

Overall, this work has developed and demonstrated a machine learning pipeline in response to the MagNet Challenge 2023. An LSTM + FNN structure is applied as the core machine learning approach. Several techniques are applied in the data pro-processing stage to enhance the model accuracy, reduce computation load and minimize bias/overfitting. For example, based on domain knowledge, a random shift/flip operation is applied on the B waveforms as pre-processing to reduce the sensitivity on the sequence of the waveform data being fed into the LSTM step for feature extraction. A special data standardization method to align the output data is applied in the fine-tuning stage to minimize the modification to the base model while cope with the small training set.

Excellent accuracy has been achieved for the 10 known materials with large-scale datasets. To tackle the challenges caused by the small datasets for the five unknown materials, special standardization and data imputation techniques are applied.

Further improvements are being tested to further balance the model performance and size, such as data imputation and model distilling methods.

REFERENCES

- [1] K. Venkatachalam, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate prediction of ferrite core loss with non-sinusoidal waveforms using only Steinmetz parameters," in Proc. IEEE Workshop Comput. Power Electron., 2002, pp. 36–41.
- [2] C. R. Sullivan, "High frequency core and winding loss modeling," in Proc. IEEE Int. Electric Mach. Drives Conf., 2013, pp. 1482–1499.
- [3] J. Li, T. Abdallah, and C. Sullivan, "Improved calculation of core loss with nonsinusoidal waveforms," in Proc. Conf. Rec. IEEE Ind. Appl. Conf. 36th IAS Annu. Meeting, vol. 4, 2001, pp. 2203–2210.
- [4] J. Wang, K. J. Dagan, X. Yuan, W. Wang and P. H. Mellor, "A practical approach for core loss estimation of a high-current gapped in-ductor in PWM converters with a user-friendly loss map," in *IEEE Transactions on Power Electronics*, vol. 34, no. 6, pp. 5697–5710, June 2019.
- [5] S. Iyasu, T. Shimizu and K. Ishii, "A novel iron loss calculation method on power converters based on dynamic minor loop," 2005 European Conference on Power Electronics and Applications, Dresden, Germany, 2005, pp. 9 pp.-P.10.
- [6] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [7] H. Li et al., "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," in *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15829–15853, Dec. 2023, doi: 10.1109/TPEL.2023.3309232.
- [8] N. Rasekh, J. Wang and X. Yuan, "Artificial Neural Network Aided Loss Maps for Inductors and Transformers," in *IEEE Open Journal of Power Electronics*, vol. 3, pp. 886–898, 2022, doi: 10.1109/OJP_EL.2022.3223936.
- [9] A. Vaswani et al., "Attention Is All You Need," *arXiv.org*, Dec. 05, 2017. <https://arxiv.org/abs/1706.03762>
- [10] X. Li and X. Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2015, pp. 4520–4524.
- [11] H. Li et al., "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," in *IEEE Transactions on Power Electronics*, 2023, pp. 15829–15853.

MagNet Challenge 2023 Report

Xinyu Liu, Chaoying Mei, Rui Zhao, Gaoyuan Wu, Hao Wu

Abstract—This report is proposed by the team from Fuzhou University, which introduces the team’s work in this competition. First, validation results on the final given data are provided including the accuracy and model size. Then, the method we used for core loss modeling is described. The network structure, fine-tuning strategy, and training tricks are discussed. Moreover, we record some attempts and observations during the experiment, which may provide potential insights.

I. RESULTS

Results on the “2023 MagNet Challenge Testing Data” are reported in this section including accuracy and model size. The accuracy is obtained on the given training data, which is further split to secondary training set and validation set. The model size is indicated by the trainable parameters in the network, which is counted by public tools and is further verified by manual calculation.

Notice: The accuracy in this report does not represent our final test results in 2023 MagNet Challenge. It only reflects the model performance on the validation set partitioned from the given training data.

A. Experimental Setup

Experiments are run on a cloud sever with Intel(R) Xeon(R) Platinum 8255C CPU, RTX 3080 GPU, Ubuntu 20.04, and CUDA 11.8. The deep learning framework is Pytorch 2.0.0 with Pytorch lightning 2.1.2 using Python 3.8. In order to ensure the reproducibility, we make the following settings in our code:

```

1 lightning.pytorch.seed_everything(666)
2 torch.backends.cudnn.benchmark=False
3 torch.use_deterministic_algorithms(True)

```

We also apply another local workstation to verify that the experiment results are reproducible. The environment of this workstation is: Intel(R) Xeon(R) w5-2465X CPU, NVIDIA GeForece RTX 4090 GPU, Windows 10 22H2, and CUDA 12.1. The deep learning framework is Pytorch 2.1.2 with Pytorch lightning 2.1.2 using Python 3.8. The validation results are consistent in both environment.

Training data in “2023 MagNet Challenge Testing Data” are further split into training subset (90%) and validation subset (10%). For data pre-processing, the frequency f and core loss P are transformed using natural logarithm function. Moreover, all the inputs and outputs are scaled by Z-score normalization. The normalization parameters include mean and variance that are calculated from all training samples (i.e., 100% of the given training data).

Xinyu Liu, Chaoying Mei, Rui Zhao, Gaoyuan Wu, and Hao Wu are with the College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: xinyu3307@163.com; 184278164@qq.com; 1131495139@qq.com; 1834237464@qq.com; 2593303519@qq.com).

B. Accuracy

Error distributions for 5 testing materials are depicted in Fig. 1. For the appointed evaluation metric in this challenge, i.e., 95th percentile error, our model achieves 3.76%, 2.04%, 2.76%, 7.58%, and 4.15% for materials A, B, C, D, and E, respectively. A common sense finding is that more data leads to higher accuracy (e.g., compare material B and D). Another observation is that if the gap between 95-Prct and Max is too large, it may mean that there are invalid samples included in the dataset (e.g., material E).

C. Model size

Five models are trained using the same network structure for five test materials. The trainable parameter number in the network is 8914. Details of the parameters counting are shown in TABLE IV, which are counted by using the Python package “torchinfo”¹. We also calculate the parameter number by hand according to the network structure and the input data size. The input size of the $B(t)$ field, frequency f , and temperature T are (1024,1), (1,), and (1,), respectively. The output P is a scalar with size (1,), which refers to the predicted core loss.

TABLE I
PARAMETERS COUNTING OF THE NETWORK.

Layer	Input size	Output size	Parameters ^a
Projector B			
Linear (Tanh ^b)	(1024,1)	(1024,24)	48
Linear	(1024,24)	(1024,24)	600
Transformer Encoder			
MHSA	(1024,24)	(1024,24)	2400
FFN	(1024,24)	(1024,24)	1984
LN	-	-	96
Projector Fusion			
Linear (Tanh)	(1024,26)	(1024,40)	1080
Linear (Tanh)	(1024,40)	(1024,40)	1640
Linear	(1024,40)	(1024,1)	41
Regression Head			
Linear	1024	1	1025

^aTotal parameters: 8914.

^bThe activation function of this Linear layer is Tanh.

It can be seen that the Transformer Encoder is the major part of the network that occupies nearly half of the parameters. This is reasonable since the excitation flux density $B(t)$ is the main input of the core loss model. It has a long sequence of sample points containing the physical information of flux density, dc-bias, waveform shape, etc. This information can only be extracted for subsequent analysis by a sufficiently powerful feature extractor. The parameter number N_{Encoder} of Transformer Encoder can be calculated by:

$$N_{\text{Encoder}} = N_{\text{MHSA}} + N_{\text{FFN}} + N_{\text{LN}} \quad (1)$$

¹[Online]. Available: <https://pypi.org/project/torchinfo/>

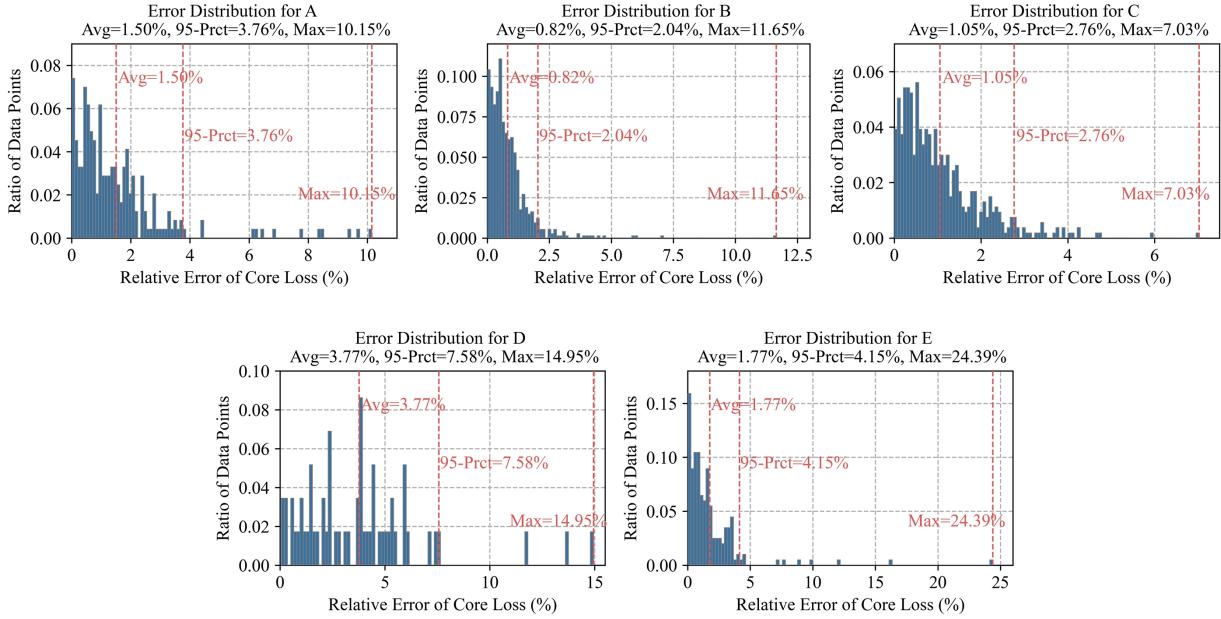


Fig. 1. Error distributions for 5 materials in “2023 MagNet Challenge Testing Data”. Results are obtained from the self-separation sets using the training data, in which 90% for training and 10% for validation.

where N_{MHSA} , N_{FFN} , and N_{LN} denote the parameter numbers of Multi-head self-attention (MHSA), Feed forward network (FFN), and Layer normalization (LN), respectively. Parameters in MHSA are constructed by the weight matrices of $W^Q \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$, $W^K \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$, $W^V \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$, $W^O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$. Moreover, each weight matrix has its bias with d_{model} parameters. The d_{model} refers to the number of expected features in the input for the Transformer Encoder, which is 24 in this report. Then, the parameter number N_{MHSA} can be computed by:

$$\begin{aligned} N_{\text{MHSA}} &= N_{W^Q} + N_{W^K} + N_{W^V} + N_{W^O} \\ &= 4 * (d_{\text{model}}^2 + d_{\text{model}}) \\ &= 4 * (24^2 + 24) \\ &= 2400 \end{aligned} \quad (2)$$

The Transformer Encoder has two FFN in which a FNN is composed of the Linear layers:

$$\text{FFN}(x) = \sigma(xW_1 + b_1)W_2 + b_2 \quad (3)$$

where σ refers to the activation function, $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{FFN}}}$ and $W_2 \in \mathbb{R}^{d_{\text{FFN}} \times d_{\text{model}}}$ are weights of the Linear layers. The $b_1 \in \mathbb{R}^{d_{\text{FFN}}}$ and $b_2 \in \mathbb{R}^{d_{\text{model}}}$ are biases. In this report, the d_{FFN} is 40. Therefore, parameters are counted by:

$$\begin{aligned} N_{\text{FFN}} &= N_{W_1} + N_{W_2} + N_{b_1} + N_{b_2} \\ &= 2 * d_{\text{model}} * d_{\text{FFN}} + d_{\text{FFN}} + d_{\text{model}} \\ &= 2 * 24 * 40 + 24 + 40 \\ &= 1984 \end{aligned} \quad (4)$$

The LN has two learnable affine transform parameters γ and β . Each of them has the same dimension as the last dimension of the input data (i.e., d_{model}). Therefore, the number of trainable parameters in LN is equal to $2 * d_{\text{model}}$. Considering

that a Transformer Encoder has two LN, the parameter number N_{LN} can be computed by:

$$\begin{aligned} N_{\text{LN}} &= 2 * 2 * d_{\text{model}} \\ &= 2 * 2 * 24 = 96 \end{aligned} \quad (5)$$

Projector Fusion is another major part in the network, which aggregates information of excitation flux density $B(t)$, frequency f , and temperature T . Its parameter number N_{ProjF} is the summary of three Linear layers:

$$N_{\text{ProjF}} = \sum_i N_{\text{L}}^i, i \in \{1, 2, 3\} \quad (6)$$

where N_{L}^i refers to the parameter number of the i -th Linear layer, it can simply be computed in terms of dimensional transformations:

$$N_{\text{L}} = d_{\text{in}} * d_{\text{out}} + d_{\text{out}} \quad (7)$$

where d_{in} and d_{out} are feature dimensions of Linear layer input and output, respectively. Then, the parameter number of each Linear layer can be calculated by:

$$\begin{aligned} N_{\text{L}}^1 &= 26 * 40 + 40 = 1080 \\ N_{\text{L}}^2 &= 40 * 40 + 40 = 1640 \\ N_{\text{L}}^3 &= 40 * 1 + 1 = 41 \end{aligned} \quad (8)$$

Projector B and Regression Head are both constructed by the Linear layer. Their parameter numbers are calculated in the same way as the Projector Fusion.

II. METHODOLOGY

Fig. 6 provides the overview of our core loss modeling method. The main idea of this method is transfer learning that reuses knowledge from previous task by using pre-training and fine-tuning approaches. It has been adopted in core loss

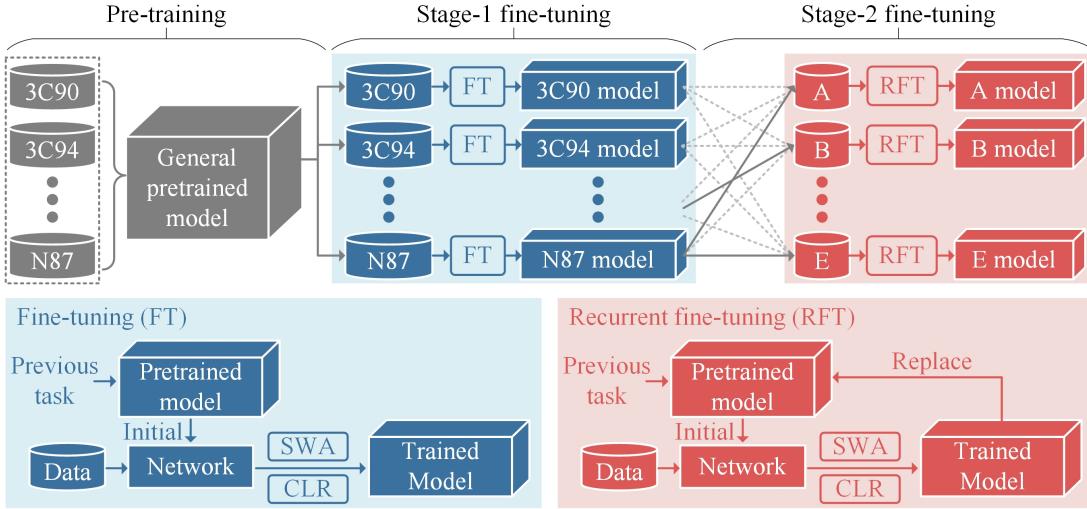


Fig. 2. Overview of our core loss modeling method based on a multi-stage fine-tuning strategy. In the pre-training, data from all 10 materials are used to train a general pre-trained model. This model is then employed in the stage-1 fine-tuning to train a specific model for each material, using data exclusively from that particular material. In stage-2 fine-tuning, stage-1 models are used as pre-trained model to train models for 5 unseen materials, respectively. Finally, we obtain 10 models for each material, but only the best model for each material is retained. SWA: Stochastic weight averaging. CLR: Cyclical learning rate.

modeling and has been proven to be effective. Based on the previous works, we introduce a number of modifications and tools to further enhance the effectiveness of the transfer-learning-based approach, which are described in this section.

A. Sequence-to-Scalar Transformer

In the tutorial provided by MagNet Challenge, the model is constructed by the sequence-to-sequence Transformer. However, in our experiments, a sequence-to-scalar model achieves a much better performance, e.g., 17.37% to 6.72% on 3F4 (For reference only). An intuitive interpretation is that predicting one data point is much easier than predicting multiple data points. As the core loss is calculated through the B - H loop in the sequence-to-sequence model, the accumulation of prediction errors from multiple data points can result in a degradation of performance.

The network structure in our method is shown in Fig. 6. Details of the network can be seen in TABLE IV. The input $B(t)$ is embedded by the Projector B and is then processed by a Transformer Encoder to extract deep abstracted features. Subsequently, the output features are aggregated with two other inputs, i.e., frequency f and temperature T . Then, the aggregated features are further processed by the Projection Fusion to mine intrinsic correlation of different kinds of inputs. Finally, the core loss P is predicted by the Regression Head.

B. Multi-stage Fine-tuning Strategy

Compared to typical fine-tuning strategy that only fine-tunes once on the target task, we apply a multi-stage fine-tuning strategy. In the pre-training (stage-0), all 10 materials are aggregated to train a General pre-trained model (GPM). And the network is randomly initialized. In the stage-1 fine-tuning, each material from 10 known materials is trained using its data. And the network is initialized from the GPM. In the stage-2 fine-tuning, each material from 5 unknown materials

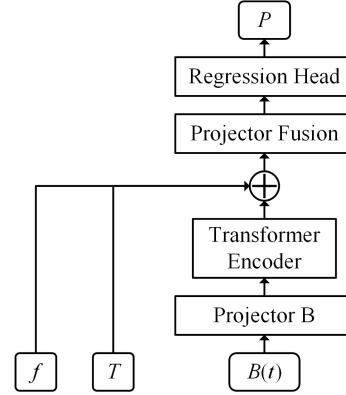


Fig. 3. Network structure of the sequence-to-scalar Transformer.

is trained using its data. And the network is initialized from the model obtained in stage-1. Finally, 10 models are trained for each material while the best one is kept.

1) *Why not use GPM in stage-2?*: We find that using GPM even worse than random initialization. A possible reason is that for each test material, there are both similar and dissimilar materials in the 10 known materials. Knowledge of dissimilar materials would affect the process of transfer learning thus leading to performance degradation.

2) *Why use GPM in stage-1?*: The impact of dissimilar materials for stage-1 is relatively small. This is because the knowledge of a known material is already included in the GPM. Thus, applying GPM is better than training from scratch. However, in our recent experiments for 10 known materials, it is sometimes better to use models from similar materials as pre-trained models. This result further validates that knowledge of dissimilar materials is likely to lead to worse transfer learning.

3) *Star material N87*.: In stage-2 fine-tuning, pre-trained models from 10 known materials show varying effectiveness,

as can be seen in TABLE II. An interesting observation is that N87 model achieves promising results among all test materials. We speculate that there may be some materials that represent the generic core loss modeling.

TABLE II
RESULTS OF STAGE-2 FINE-TUNING (TRAIN:VAL=8:2).

Material	A	B	C	D	E
Scratch	5.36	2.61	3.71	14.21	7.32
3C90	4.73	2.30	3.09	13.03	5.67
3C94	4.86	2.41	3.21	13.95	5.47
3E6	4.71	2.55	3.78	17.93	6.03
3F4	6.14	2.82	3.55	15.71	6.32
77	5.20	2.61	3.29	14.05	6.03
78	5.01	2.65	3.34	14.88	5.42
N27	5.51	2.69	3.47	13.16	5.94
N30	5.37	2.34	3.56	21.18	6.25
N49	5.43	2.31	3.55	18.68	6.61
N87	4.51	2.27	3.10	12.07	5.13

4) *A possible way to foundational core loss model.*: Solving all problems with one foundational model is a current research goal in many domains such as CV and NLP. The results in transfer learning of core loss modeling show that a better pre-trained model can achieve better performance. This suggests that there is some general knowledge in deep learning models that can be used for other new materials. This also provides a possible way to build a foundational model for core loss modeling, which can be adapted to different materials without any adjustments, just like large language model GPT. However, before we can achieve this ambitious goal, we first need to build a foundational pre-trained model: more diverse data and better pre-training method.

C. Tricks for Model Training

1) *Efficient Coding Framework for Deep Learning*: Coding is important in our research pipeline. We refactor the deep learning code with Pytorch Lightning, which greatly accelerated our research process. We do the research, Pytorch Lightning does the engineering.

2) *Mixed Precision*: By conducting operations in half-precision format while keeping minimum information in single-precision to maintain as much information as possible in crucial areas of the network, mixed precision training delivers significant computational speedup without accuracy loss.

3) *Cyclical Learning Rate*: We use Cyclical learning rate (CLR) policy instead of warm up. The CLR policy cycles the learning rate between two boundaries with a constant frequency, which eliminates the need to perform numerous experiments to find the best values and schedule with essentially no additional computation.

4) *Stochastic Weight Averaging*: Stochastic Weight Averaging (SWA) is adopted after CLR training, which achieves significantly better generalization at virtually no additional cost. SWA performs an equal average of the weights traversed by a stochastic optimizer (Adam in this report) with a modified learning rate policy. In our understanding, SWA's role is similar to the model ensemble that can aggregate multiple learners to build a strong learner.

5) *Recurrent Fine-tuning*: Based on CLR and SWA, we propose a Recurrent fine-tuning (RFT) approach which consists of initial fine-tuning and continuous fine-tuning. The training log of RFT is depicted in Fig. 4. In the initial fine-tuning, the network is initialized by using the model from previous task, e.g., N87 model obtained from stage-1 fine-tuning. In the continuous fine-tuning, the well-trained model from initial fine-tuning is used as pre-trained model. In our understanding, RFT equates to increasing the training epoch, but has two differences. First, CLR ensures that the model does not always fall into a local optimum. Second, SWA is applied twice in the entire training period, which improves pre-trained model quality for latter half training process.

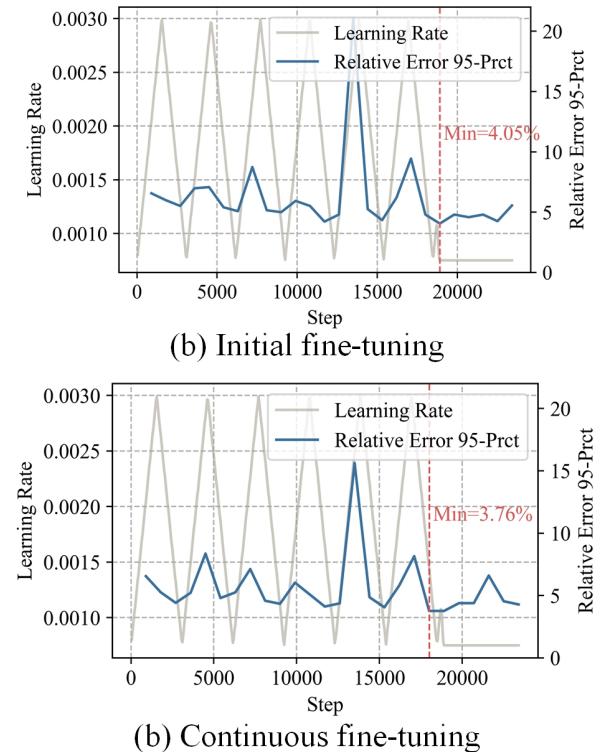


Fig. 4. Training log of material A using RFT.

Results of RFT are shown in TABLE II. In materials A, B, and C, RFT achieves a better performance than that of fine-tuning directly using a longer training time. However, RFT may encounter overfitting in materials with small sample.

TABLE III
RESULTS OF RECURRENT FINE-TUNING. (TRAIN:VAL=9:1)

	A	B	C	D	E
Scratch	4.90	2.28	3.61	14.45	7.61
Fine-tuning	4.05	2.17	2.93	7.58	4.74
Fine-tuning (long)	4.21	2.13	2.93	7.58	4.15
Recurrent fine-tuning	3.76	2.04	2.76	7.79	4.29

III. DISCUSSION OF SOME ATTEMPTS AND OBSERVATIONS

A. Extra Data from Self-measurement

The performance of the data-driven method is highly dependent on the data fed into the model. Deep learning, a data-

hungry machine learning paradigm, demands more from the amount of training data. It's a natural idea to use extra data for better performance. Therefore, we construct a measurement system to obtain more samples. However, after obtaining very unsatisfactory results (6.69% to 14.07% on 3F4) and considering the expensive labor cost, we abandon this approach.

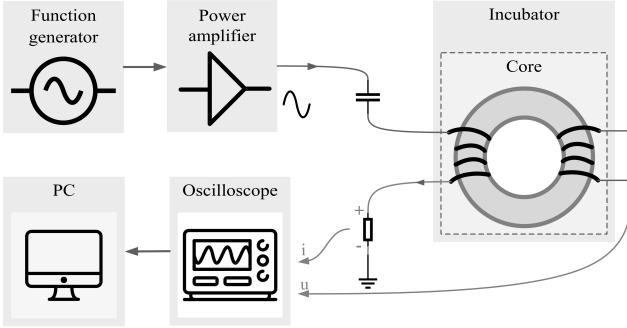


Fig. 5. Overview of our measurement system.

1) Measuring Environment: The overview of our measurement system is shown in Fig. 5. The core loss is measured by using the two-winding method. Sinusoidal excitation with different frequencies and amplitudes is provided by a function signal generator (Agilent 33522A) that is connected to the power amplifier (NF HSA4014). The voltage and output power of the excitation thus can be amplified and then be applied to the measured magnetic component. The operating temperature of the magnetic component is controlled by an incubator. Subsequently, the voltage waveform and current waveform are acquired using an oscilloscope (HD4096). Finally, by processing the acquired data, the magnetic flux density curve, magnetic field intensity curve, and magnetic core loss can be obtained. The measuring devices can be seen in Fig. 6.



Fig. 6. Devices for data acquisition.

For the measured material, we choose 3F4 that has the minimum training data among 10 materials provided in pretest phase. Details of the measuring environment are depicted in TABLE IV. Notice that the frequency interval is 25 KHz and all curves of B and H are resampled to 1024 sampling points. Finally, 662 samples are obtained and converted to MagNet format. These extra data will be packaged into our commit files for possible research purposes.

2) Analysis of Performance Degradation: Possible reasons for performance degradation are: 1) Differences in measurement environments. Although we try to emulate the measurement environment of MagNet as much as possible, due to the limited resources, it is difficult to achieve complete

TABLE IV
DETAILS OF THE MEASURING ENVIRONMENT.

Property	Information
Temperature	25°C, 50°C, 75°C, 90°C
Frequency	100 KHz ~ 500 KHz
Excitation	Sinusoidal wave
Sampling interval	4×10^{-9} s
Sampling point	500 ~ 2500

consistency. In particular, the frequency and excitation cannot be aligned with that of the MagNet data. 2) Differences in pre-process and post-process of the measured data. A standard publicly available guideline or toolkit may narrow the differences mentioned above. 3) The amount of extra data is still too small that can not cover more working conditions. One solution to alleviate this problem is to improve the intelligence of our measurement procedure. 4) Excitation is sinusoidal wave only, which leads to a significant bias in the model.

B. Other Attempts and Observations

1) Inconsistency of Training, Validation, and Testing: In deep learning coding, there are many factors that would affect the model performance. Here we share some of the factors we encountered: 1) Mixed Precision. If mixed precision is used during training, there may be subtle differences in the results when testing with full float precision. 2) Batch size. Using different batch size for training and testing produces different results even if no batch-related modules are used. We still haven't figured out why. 3) Method in "np.percentile". Different computing methods lead to different 95-Prct results.

2) Frozen Layers: Freeze some layers (usually shallow layers) in pre-trained model and fine-tune the rest is a common trick for a deep learning application. However, this trick did not work well in our experiments (2.85% for fine-tune all layers and 3.34% for freeze Transformer Encoder on material B), probably because the pre-trained model is not powerful enough.

3) Projection of T and f: We project (or embed) T and f before Projector Fusion in an attempt to allow the model to better mine knowledge of the measuring environment. However, performance get worse (14.21% to 15.85% on material D) while model size is larger (8.9K to 11K).

4) Data Normalization: The Z-score normalization parameters are calculated using a group of samples (The more, the better), and they are material-specific (one-to-one). If the test material type is unknown, the normalization can not be applied. A general normalization parameter or a material-independent normalization approach may solve this problem.

5) Measurement Error: The equation to obtain B curve assumes a uniform distribution of magnetic induction intensity within the core. However, the distribution of magnetic induction intensity varies at different positions. Under large excitation, due to local saturation, the actual loss may exceed the computed value.

MAG-ViT: A ViT Backbone for Magnetic Material Modeling

Rui Zhang, Hao Song, Lie Zhang, Yibo Lu, and Lijun Hang, *Senior Member, IEEE*,

Abstract—Classical magnetic core loss modeling methods have limitations in terms of accuracy and comprehensiveness. Existing deep learning modeling methods also suffer from limited data size and overly simple models. Vision transformers (ViT) have shown promise in various vision tasks. In this study, we proposed MAG-ViT, which uses ViT as backbone for magnetic material modeling. MAG-ViT extends the magnetic material model to deeper dimensions in a more sensible way while using less memory through patch embedding, class token and learnable positional embeddings method. The open source MAGNET dataset was utilized for training and evaluating of MAG-ViT. The experimental results show that MAG-ViT performs well on the core loss prediction task and it ranks 8th in pretest results. Our test scripts and models are available at <https://github.com/moeKedama/dg-magnet-test-script>.

Index Terms—deep learning, power magnetics, core loss, hysteresis loop, vision transformer (ViT).

I. INTRODUCTION

MAGNETIC components are an important part of power electronics systems for energy buffering. The design of magnetic components relies heavily on classical loss models for magnetic materials and material datasheets tailored to specific operating conditions, such as waveforms, frequency, and temperature. Classical magnetic core loss models, including the *Steinmetz equation*, *iGSE* [1], rely on empirical simplifications and physical approximations. Despite numerous upgrades since the introduction of the Steinmetz equation in 1890, these models still have limitations regarding accuracy and comprehensiveness [2]–[4]. These models may introduce significant errors by neglecting the intricate behavior of magnetic materials and their variations under diverse operating conditions. However, manufacturers' datasheets, while providing more accurate data, only cover a limited range of operating conditions. Acquiring data for all conceivable operating conditions would require significant testing efforts and resources. Obtaining an accurate and general model with a small amount of experimental data is the crux of the problem.

In recent years, there has been an increasing interest in modeling the magnetic materials with neural networks [5]. Training neural networks using both experimental and simulation data enables the construction of models for the behavior of magnetic materials. These models find applications in the design phase, aiding in the optimization and enhancement of magnetic components. Neural-network-based methods, which operate independently of physical models, demonstrate superior generalization capabilities. They excel in capturing the

This paper was produced by the team of Hangzhou Dianzi University.
Our Github repository: <https://github.com/moeKedama/dg-magnet-test-script>

intricacies of complex magnetic material behavior and various operating conditions. In addition, these methods provide better coverage of the overall data distribution compared to classical datasheets. Studies have demonstrated that neural-network-based methods incur lower computational overhead compared to Preisach model-based methods [6]. Nonetheless, many existing methods are constrained by small dataset sizes, and their model structures are often too simplistic to exhibit effective generalization when meeting real-world task requirements. These challenges can be addressed by leveraging the newly developed open-source dataset, MAGNET [7]–[9]. Leveraging the MAGNET dataset facilitates greater comparability of results across related studies compared to development on disparate datasets with unknown data quality, limited size, and singular operational situations.

The Transformer has consistently stood as the state-of-the-art approach for numerous sequence-to-sequence tasks since its inception [10]. It utilizes a self-attention mechanism to capture temporal relationships between input and output sequences. Despite the limitations of the classical Transformer model in handling inputs, such as insufficient processing of local information and fixed size requirements, Vision Transformer (ViT) addresses these challenges [11]. ViT closely adheres to the original Transformer while introducing a simple yet highly extensible structure. It demonstrates out-of-the-box efficiency, providing an improved solution for handling large inputs and capturing essential local details.

We propose a ViT-based modeling method for hysteresis loop and core loss modeling, called MAG-ViT. MAG-ViT exhibits a significant improvement over Transformer-based methods, boasting reduced memory consumption, enhanced modeling accuracy, and greater scalability.

The paper is structured as follows: Section II details the specific modeling approach employed by MAG-ViT; Section III outlines the experimental results obtained with MAG-ViT on the MAGNET dataset; Section IV concludes this paper by summarizing the presented work and suggesting directions for future exploration.

II. MAG-ViT

ViT operates by segmenting the input into fixed-size patches and embedding them into vectors. It also leverages the self-attention mechanism to capture global and local relationships within the input. In order to apply ViT to magnetic material modeling, it is crucial to preprocess the magnetic material input. The model structure is also adjusted to fit the specific modeling requirements. Here, we introduce MAG-ViT, which

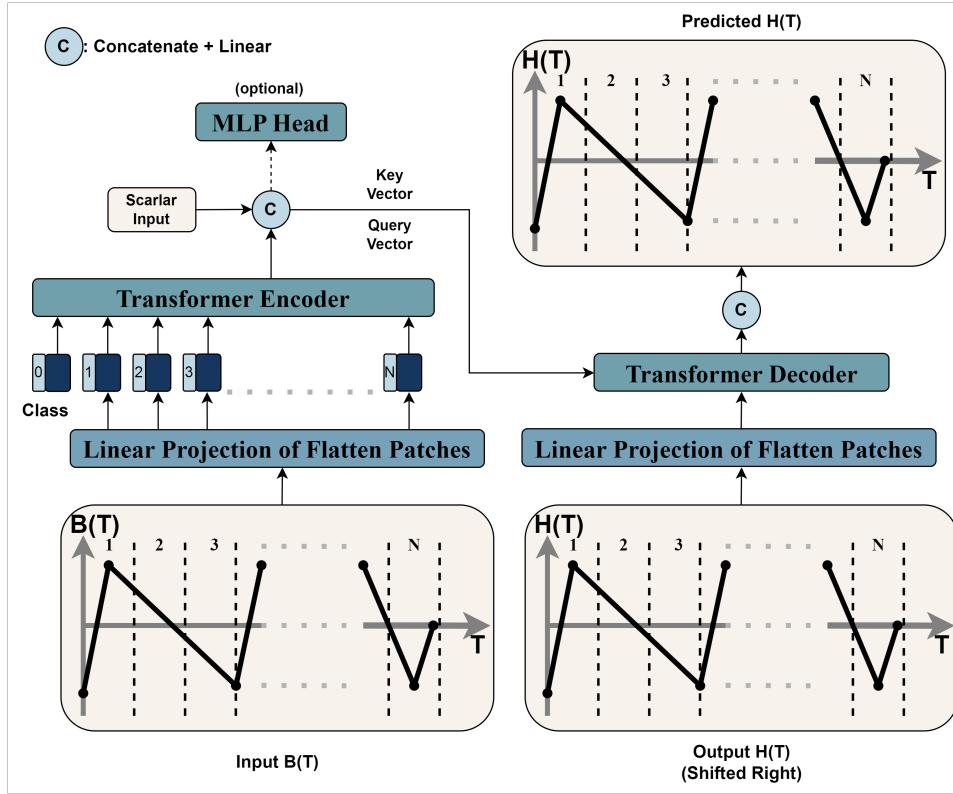


Fig. 1. Model overview and data flow of MAG-ViT.

incorporates 3 major improvements over the transformer approach.

1) Patch Embedding: The input to the general Transformer comprises a 1D sequence of token embeddings. To manage the flux density sequence $B(t) \in R$, we initially employ a Multi-Layer Perceptron (MLP) to linearly project $B(t)$ to C dimensions. Subsequently, the projected $B(t)$ sequence, with a C -dimensional latent vector size, undergoes transformation into a sequence of flattened 2-D patches with $x_p \in N*(p^2*C)$. In this context, the input sequence can be conceptualized as a sequence of flattened 2-D blocks. The total count of flattened 2-D blocks is $N = R/P^2$ and the dimension of each block is $(P^2 * C)$, where P denotes the size of the block.

2) Class Token: A BERT-like class token, represented by the learnable embedding x_{class} is introduced to the sequence of embedded patches. x_{class} shares the same dimensions as the input sequence $B(t)$, resulting in an output of $N + 1$ vectors. Functioning akin to the average pooling layer in a Convolutional Neural Network (CNN), the class token integrates information across the model input tokens. Similar to the role of the average pooling layer in CNN, which is utilized for synthesizing information from input tokens, the class token plays a crucial role. After interacting with each patch for information, it enables the model to learn specific classification information. Additionally, within the self-attention mechanism, the interaction between the class token and various patches is contingent on the degree of interaction. This approach allows the model to discern which patches exert influence on the final classification results. Moreover,

it enables the model to determine the specific degree of that influence, thereby enhancing the interpretability of the model.

3) Learnable Positional Embeddings: Positional embeddings are imperative for Transformers, given that altering the token sequence order has no impact on the outcome. The absence of positional embeddings increases the learning cost as the model, lacking positional information, resorts to relying solely on the semantics of the patches. As an enhancement, standard learnable 1D positional embeddings, denoted as E_{pos} are incorporated as a substitute for absolute positional encodings. The resulting sequence of embedding vectors, inclusive of positional information, is then fed into the encoder. This modification contributes to the model's ability to understand and utilize the positional relationships within the sequence.

Equation (1) delineates the complete process of preprocessing $B(t)$ and $H(t)$ sequence

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos}, \quad (1)$$

where $x_p \in (x_p^1, x_p^2, \dots, x_p^N)$ denotes the sequence patch, and E denotes the linearly projection operation.

The network structure of MAG-ViT is depicted in Fig. 1. Initially, the data points of each time step in the input sequence $B(t)$ undergo transformation into a representation of patch embeddings with dimension D and length $N+1$. This transformation is achieved through the linear projection of the flattened patches block. The resulting representation encompasses class tokens and learnable positional embeddings. Subsequently, the generated patch embeddings are inputted into the Transformer Encoder block. This block analyzes and captures the temporal

TABLE I
MAXIMUM PATH LENGTHS, PER-LAYER COMPLEXITY AND MINIMUM NUMBER OF SEQUENTIAL OPERATIONS FOR DIFFERENT LAYER TYPES

Layer Type	Complexity per Layer	Sequential Operations	Maximum Path Length
Convolution	$O(k \cdot n \cdot d^2)$	$O(1)$	$O(\log_k n)$
Recurrent	$O(n \cdot d^2)$	$O(n)$	$O(n)$
Self-Attention (Transformer)	$O(n^2 \cdot d)$	$O(1)$	$O(1)$
Patched Self-Attention (ViT)	$O(\frac{n^2}{patch_length} \cdot d)$	$O(1)$	$O(\frac{n}{patch_length})$

dependencies within the input sequences, yielding sequence-informative query vectors and key vectors.

The query vector and key vector are concatenated with scalar inputs (i.e., temperature, frequency, peak-to-peak value of $B(t)$). The concatenated vectors are then mapped back to their original dimensions through a linear projection layer. These vectors are further transformed to the dimensions of the original query vector and key vector, serving as new query vector and key vector inputs to the Transformer Decoder. Also, an optional MLP Head can be set up here for classification or regression tasks. The Transformer Decoder is responsible for reconstructing the output sequence.

For the implementation of output sequence reconstruction, the Transformer Decoder requires a value vector. During the network training phase, the value vector is generated using the sequence $H(t)$ as input. In the network testing phase, the value vector is cyclically generated, with the predicted output sequence initialized to 0 as input. Subsequent to further processing by the Transformer Decoder, the model produces a patched output $H(t)$ sequence with dimensions D and length $N + 1$.

The initial N patches of the sequence are extracted, and their dimensions are transformed from D to p through a linear projection. Each D -dimensional patch contains information about p original data points. The final prediction $H(t)$ is obtained by concatenating these patched sequences in order.

In contrast to the down-sampling methods of example, MAG-ViT aims to retain maximal information at high frequencies while maintaining an acceptable complexity level. It achieves this by encapsulating high-frequency features within patches and reflecting low-frequency features between patches. This model possesses a higher performance upper bound compared to down-sampling methods, albeit at the cost of increased training expenses.

It is essential to highlight that the inductive bias of ViT is notably smaller compared to Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs exhibit properties such as locality, spatial invariance, translational equivariance, whereas RNNs possess properties like sequential, time invariance. In the ViT-based approach, only the MLP layers are local and translationally equivariant, while the self-attention layers operate globally. The positional embeddings must be trained to be sufficiently expressive to handle relationships between entities effectively. This requirement also implies that the dataset used for training needs to attain a certain size.

Table I is maximum path lengths, per-layer complexity and minimum number of sequential operations for different layer

TABLE II
THE NUMBERS OF PARAMETERS IN THE MODEL

Model	Number of Parameters
Simple	2,396,048
Full(With MLP Head)	2,528,913

types, where n is the sequence length, d is the representation dimension and k is the kernel size of convolutions. Compared to Self-Attention, Patched Self-Attention has a smaller per-layer complexity.

III. EXPERIMENTS

We evaluated the proposed MAG-ViT in the hysteresis loop and core loss modelling tasks. Before presenting these results, we outline the main experimental setup below. Additional details can be found in our GitHub repository.

A. Experimental Setup

Datasets. For class-conditional learning, we consider the MAGNET dataset [magnet citation], which contains 10 different ferrite materials across a wide MAGNET dataset contains 4 fields which are flux density waveform $B(t)$, field strength waveform $H(t)$, fundamental frequency f , and temperature T . In the preprocessing stage, normalization is applied to all fields within the MAGNET dataset. The normalization parameters are preserved for subsequent testing and inference.

MAG-ViT configuration. The dimension of the linear projection output, denoted as dim_val , is set to 32. For the original input $B(t)$ with a length of 1024, the shape after linear projection becomes (1024, 8). To maintain a balance between the length of the patch embedding sequence and the total number of tokens in each patch, the patch size utilized by MAG-ViT is configured as (8, 32). Consequently, for the input $B(t)$ after linear projection, the corresponding output patch embedding shape is (128, 256). In this context, every 8 original data points projected to 32 dimensions form a patch, and each patch comprises 256 tokens (represented by $embed_dim$ in the configuration). Concerning the transformer configuration, the number of neurons in the linear layer of both the transformer encoder and decoder is set to 256. The number of attention heads is 8, and there is a single stacked encoder layer in both the encoder and decoder. Following this configuration, the total number of parameters in the models are shown in TABLE II. Both 2 models are under the 10MB parameter limitations.

Training. We employ the Adam optimizer with an initialized learning rate of 0.0001 for all materials. All models undergo

TABLE III
CORE LOSS PREDICTION RELATIVE ERRORS OF THE MAG-ViT

Model	AVG	95-Prct	99-Prct	Max
3C90	2.052%	6.38%	11%	29.66%
3C94	1.834%	5.64%	10.4%	44.36%
3E6	0.6745%	1.56%	2.24%	7.097%
3F4	3.673%	11.4%	21.1%	50.26%
77	1.988%	4.77%	8.19%	19.74%
78	1.917%	5.65%	9.78%	28.83%
N27	1.668%	5.33%	8.95%	27.24%
N30	0.6371%	1.6%	2.58%	19.39%
N49	3.059%	10.4%	18.5%	57.77%
N87	1.58%	4.77%	8.44%	34.98%
average	1.908%	5.75%	10.118%	31.9327%

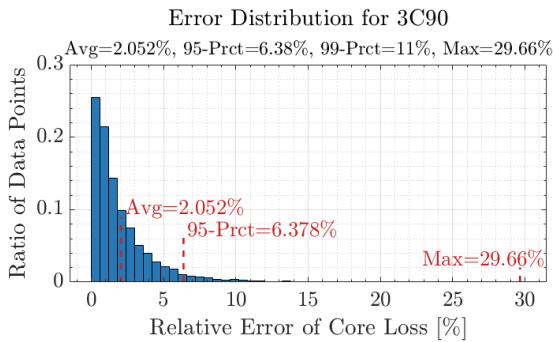


Fig. 2. Core Loss Prediction Relative Errors of Material 3C90

training for 4000 epochs, with a manual reduction of the learning rate to 0.1x at 2700 and 3600 epochs. The training process occurs at a rate of approximately 200 epochs per 5 minutes on a PC equipped with a 5.80GHz i9-13900k CPU, 128GB DDR5 4200MHz RAM, and an NVIDIA RTX A6000 GPU. The model consumes approximately 0.6GB of VRAM with a batch size of 32. MAG-ViT employs mean-square error (MSE) as the loss function for training. Each material is trained separately, resulting in 1 models for each material. To achieve the results presented in the Pretest Results, convergence should be at least on the order of 10e-5 to 10e-6 on the training set. Moreover, it should be at least on the order of 10e-4 to 10e-5 on the validation set, which does not participate in the training process.

B. Results on MAGNET pretest Dataset

The MAGNET pretest Dataset serves as a validation set accompanying the MAGNET dataset, encompassing 10 existing materials, each comprising 5,000 randomly sampled data points from the original database.

In the Core Loss Prediction task, the predicted core loss $P_{V_{pred}}$ can be directly computed through integrals with respect to the original input $B(t)$ and the predicted $H(t)$. Subsequently, the relative error between the predicted core loss $P_{V_{pred}}$ and the measured $P_{V_{mea}}$ is employed as an evaluation metric.

$$P_V = \frac{1}{T} \int_{B(0)}^{B(t)} H(t) dB(t) \quad (2)$$

TABLE IV
THE SUB-DOMAINS TEST OF THE MAG-ViT

Model	AVG	95-Prct	99-Prct	Max
3C90	13.07%	37.2%	144%	344.7%
3C94	12.2%	34.5%	129%	383%
3E6	5.21%	19.4%	46.7%	90.4%
3F4	17.49%	62.2%	119%	407.5%
Sub1 avg	12.0175%	38.325%	109.675%	306.4%
77	2.569%	7.18%	10.1%	27.69%
78	3.591%	10.7%	15.5%	33.49%
Sub2 avg	3.08%	8.94%	12.8%	30.59%
N27	22.35%	70.9%	113%	346.2%
N30	16.8%	54.2%	75.3%	268.9%
N49	36.26%	138%	274%	599%
N87	20.3%	62%	108%	338.8%
Sub3 avg	23.9275%	81.275%	142.575%	388.225%
average	14.994%	49.628%	103.46%	283.968%

$$\text{Relative Error} = \frac{|P_{V_{pred}} - P_{V_{mea}}|}{P_{V_{mea}}} \quad (3)$$

Table III lists the results of our study of the 10 materials in the dataset. Here, AVG represents the average relative error, 95-Prct denotes the 95th percentile relative error, 99-Prct represents the 99th percentile relative error, and Max signifies the maximum relative error.

Taking 95-Prct-Error as an example, we achieved the best results by reaching 1.56% on material 3E6. This ranked us 3rd out of all 25 teams that submitted pre-test results. On the worst-performing material, 3F4 reached 11.39% and ranked 9th. The average 95-Prct-Error for all materials combined ranked 8th.

Fig. 2 displays the details of our study on material 3C90. The majority of the data points have an error of less than 5%. This indicates that the model is generally accurate in its predictions of the core loss. However, there is a small number of data points with a much higher error. These data points may indicate that the model is not as accurate for some types of data or they may be outliers.

After evaluating the model on 10 materials, we categorized the materials into 3 sub-domains.

- sub-domain 1 : 3C90, 3C94, 3E6, 3F4.
- sub-domain 2 : 77, 78.
- sub-domain 3 : N27, N30, N49, N87.

The division of the 3 sub-domains is based on potential similarities between the 10 materials. We posit that certain weights in the network can be shared among similar materials, particularly in the encoding layers. To test this hypothesis, we trained the sub-domain data using the MAG-ViT configuration, only altering the number of stacked encoder layers from 1 to 2. The modified model has a total of 3,586,458 parameters. Each sub-domain is trained separately.

Table IV presents the results of the MAG-ViT sub-domain tests. Materials in sub-domain 2 achieved similar results to those in Table III, whereas materials in sub-domain 1 and sub-domain 3 performed poorly in the sub-domain tests. Several possible causes are considered:

1) Parameters Limitations: In the Magnet Challenge, the size of all included models was restricted to 10 MB for one material. Despite using 14.2 MB parameters in the sub-domain test, we did not achieve satisfactory results, except for sub-domain 2. We speculate that more similar materials require fewer parameters, and sub-domains with a greater number of materials necessitate more parameters. Model quantization and model distillation methods can be employed to conserve resources in the model deployment phase. The 10MB limit may be too stringent for research-stage models, especially when the model needs to encode multiple materials.

2) Data normalization methods: The current normalization method involves computing means and standard deviations across all domains or sub-domains, relying on prior knowledge. However, when there is more than one material in the training data, this method may not normalize the data effectively. As a result, the model can face challenges in convergence, particularly in cases where prior knowledge is not fully applicable.

3) Lack of inductive bias: ViT's performance is constrained on small and medium-sized datasets due to the lack of inductive bias. The Magnet dataset, comparable in size to datasets like MNIST and CIFAR, which are considered small datasets, faces limitations in this context. Larger dataset sizes are crucial for advancing future research in this domain.

IV. CONCLUSION

This work introduces MAG-ViT, a straightforward and versatile ViT-based architecture for magnetic material modeling. MAG-ViT processes sequential and scalar inputs to construct embeddings and is assessed in tasks such as Hysteresis B-H Loop Prediction and Core Loss Prediction. Experimental results indicate that MAG-ViT exhibits promising performance. We believe that MAG-ViT can offer valuable insights for future research on backbones in magnetic material modeling and contribute to endeavors such as transfer learning in this field.

ACKNOWLEDGMENTS

This paper constitutes the final submission of the Hangzhou Dianzi University team for the IEEE PELS-Google-Enphase-Princeton MagNet Challenge. The open-sourced dataset utilized in this work is sourced from the MagNet Team.

REFERENCES

- [1] J. Goodenough, "Summary of losses in magnetic materials," *IEEE Transactions on Magnetics*, vol. 38, no. 5, pp. 3398–3408, 2002.
- [2] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 964–973, 2012.
- [3] Y. Han, G. Cheung, A. Li, C. R. Sullivan, and D. J. Perreault, "Evaluation of magnetic materials for very high frequency power applications," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 425–435, 2012.
- [4] M. Chen, M. Araghchini, K. K. Afandi, J. H. Lang, C. R. Sullivan, and D. J. Perreault, "A systematic approach to modeling impedances and current distribution in planar magnetics," *IEEE Transactions on Power Electronics*, vol. 31, no. 1, pp. 560–580, 2016.
- [5] Y. LeCun, Yann Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436–444, 2015.

- [6] S. Quondam Antonio, F. Riganti Fulginei, A. Laudani, A. Faba, and E. Cardelli, "An effective neural network approach to reproduce magnetic hysteresis in electrical steel under arbitrary excitation waveforms," *Journal of Magnetism and Magnetic Materials*, vol. 528, p. 167735, 2021.
- [7] H. Li, D. Serrano, S. Wang, and M. Chen, "MagNet-AI: Neural Network as Datasheet for Magnetics Modeling and Material Recommendation," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15854–15869, Dec. 2023.
- [8] D. Serrano, H. Li, S. Wang, M. Luo, T. Guillod, C. R. Sullivan, and M. Chen, "Why MagNet: Quantifying the Complexity of Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14292–14316, Nov. 2023.
- [9] H. Li, D. Serrano, T. Guillod, S. Wang, E. Dogariu, A. Nadler, M. Luo, V. Bansal, N. Jha, Y. Chen, C. R. Sullivan, and M. Chen, "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15829–15853, Dec. 2023.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Reinforcement (ICLR)*, Vienna, Austria, 2021, pp. 1–22.

2023 MagNet Challenge: An Approach Towards Improved Core Loss Modelling

Neha Rajput*, Himanshu Bhushan Sandhibigraha†, Neeraj Agrawal§, and Vishnu Mahadeva Iyer¶,

Department of Electrical Engineering, Indian Institute of Science, Bengaluru - 560012

Email: *neharajput@iisc.ac.in, †himanshub@iisc.ac.in, §aneeraj@iisc.ac.in, ¶vishnumi@iisc.ac.in

Abstract—2023 MagNet challenge aims to develop innovative core loss modeling approaches with high accuracy and reduced complexity. The core loss of magnetic materials is governed by complex physics and depends on the material's electrical, magnetic, mechanical, thermal, and other physical properties. An evaluation of state-of-the-art empirical and data-driven approaches for core loss modeling has been done. In our approach we aim to leverage the best of both empirical and data-driven methods, thus coming up with a hybrid modeling strategy. This report summarises the key aspects of the hybrid model and summarizes our team's efforts toward developing the hybrid core loss model.

I. INTRODUCTION

Power magnetics constitutes about 50-60% of the weight and volume of state-of-the-art power converters. Design optimization of power-magnetics is crucial to enable highly efficient and power-dense converters for weight-critical systems.

Magnetics design in itself is a multidisciplinary endeavor. A designer has many choices regarding magnetic core materials, core geometries, and winding options. Typically, several design iterations are needed to arrive at an optimal design. The availability of accurate models (reluctance models, core loss models, copper loss models, thermal models, insulation aging models) can considerably speed up the design life cycle.

The focus of this work is to come up with improved methods for modeling magnetic core losses. The 2023 MagNet challenge has renewed researchers' interest in developing improved core loss modeling methods. The availability of experimental datasets has made it easier to evaluate existing models and benchmark any new models. We target a multi-pronged approach toward modeling the core losses that combines the simplicity of the empirical model and the accuracy of the data-driven approaches. Our approach includes separating the input data based on the excitations (sine, triangular and trapezoidal) and building three hybrid models i.e., we are trying to embed the nature of physics with the neural network. Our approach includes the following:

- Evaluating state-of-the-art empirical models (listed in Table I) using the experimental data set and identifying any underlying trends by analyzing the error data.
- Building hybrid neural networks i.e., modeling three separate physics-informed neural networks (PINN) for each sinusoidal, triangular, and trapezoidal excitations; training them by embedding suitable empirical core loss equations and parameters trends with respect to temperature and frequency into the training process.

- Building a pre-trained model by using the pre-existing material data (previously published 10 materials) for each excitation.
- Fine-tuning and validating the PINN models using the hyperparameters of the models for each excitation for the newly introduced five materials.

II. OUR APPROACH

In this section, the steps taken towards creating the hybrid models are explained in detail.

A. Feature Extraction

First, the single-cycle flux density waveform data is segregated into three datasets based on the excitation i.e., sinusoidal, triangular, and trapezoidal excitations. Several distinct features characterizing each waveform have been extracted and listed below for each excitation type.

- Sinusoidal flux type - peak-to-peak flux density (B_{pkpk}), maximum flux density (B_{max}), minimum flux density (B_{min}), time period (T).
- Triangular flux type - peak flux density (B_{pk}), duty ratio (d), positive and negative slopes (dB/dt)
- Trapezoidal flux type - flux densities (B_1, B_2, B_3, B_4) at the critical points (i.e. instants where there is a slope change), duty ratios (d_1, d_2, d_3, d_4).

In addition to the above quantities, the frequency (f) and temperature (T) data are directly read from the data provided by the MagNet team and have been added as features of the respective waveforms.

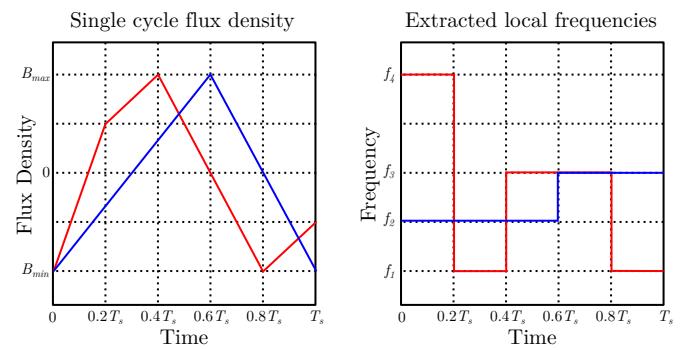


Fig. 1. Composite waveform hypothesis-based extraction of local frequencies [8].

TABLE I
LITERATURE SUMMARY

Loss Models	Highlights	No. of Parameters
Jiles-Atherton model [1]	Accurate phenomenological model describing the hysterical behavior of the magnetic materials. A thorough understanding of magnetic domain walls and other microscopic behavior is required for accurate macroscopic hysteresis loss modeling.	5
Steinmetz equation [2]	A simple power law equation-based empirical model, which is suitable for sinusoidal excitations only.	3
Modified Steinmetz equation (MSE) [3]	First empirical model to identify the effect of induction change-rate $\frac{dB}{dt}$ on core-loss and introduction of an equivalent-frequency concept.	3
Improved Generalized Steinmetz Equation (iGSE) [4]	Influence of peak-to-peak flux density on core loss rather than the instantaneous value of flux density is discussed.	3
Improved Steinmetz Equation (ISE) [5]	A composite waveform hypothesis (CWH) based model where the effect of waveform duty ratio is incorporated. It is similar to the Steinmetz equation, where frequency and flux density are the main variables.	up to 6 for triangular and up to 12 for trapezoidal excitations
MagNet core loss model [6]	A two-stage machine learning for magnetic core loss modeling. The complex behavioral attribute of magnetic material can be captured using a machine learning approach.	upto 50,000
Meta-material core loss model [7]	Introduced a new core loss model for estimating eddy current core loss based on the meta-material modeling approach of material parameters and wave equations; when combined with state-of-the-art hysteresis model can predict accurate estimation of overall core losses	11

B. Empirical Models

State-of-the-art empirical loss models that are available in the literature are evaluated.

In our approach to hybrid modeling, we have used the Improved Steinmetz equation (ISE) [5] based empirical loss equations. It reduces to the Steinmetz equation for sinusoidal data. The ISE has been extended for trapezoidal excitations by applying the composite waveform hypothesis (CWH) [8] as shown in Fig. 1.

Further, the coefficients used in the ISE based models have been made adaptive with respect to frequency and temperature rather than keeping them as constants for improved accuracy.

C. Proposed Hybrid Model

Our approach to hybrid modeling includes the following innovative aspects.

Three separate models

Three different hybrid models have been developed, one for each excitation. The hybrid modeling approach is inspired by physics-informed neural networks, wherein the empirical equations governing the core loss are incorporated into the neural network. The incorporation of the empirical loss model into the neural network is shown in Fig. 2.

Creating separate models allows the flexibility to embed separate empirical core loss functions suitable for each excitation. Table II shows the architecture of the neural network for each excitation type. It is to be noted that these three models are used for training data for a single material.

Learnable parameters

The coefficients of the empirical loss models have been made adaptive with respect to temperature (T) and frequency (f). Therefore, they are expressed as functions of T , f , and some constants. These constants are defined as the learnable

parameters of the neural networks. This enhances the capacity of the neural network to learn meaningful patterns and generalize its learning. This will help to predict the results for data that has not been encountered while training. Also, the coefficients in the empirical core loss models exhibit wide variations across different grades of ferrite materials. Thus, including these coefficients as the learnable parameters seemed like an interesting and wise option.

Also, these meaningful patterns can further be utilized to understand the core loss trends for each material, for which obtaining an empirical fit is challenging.

Customized training loss function

In our training process, a custom loss function is defined. This function ensures that the output of the neural networks, i.e., core loss of the given ferrite material, closely follows the defined empirical equation. The custom loss, $Loss_{custom}$ is defined as:

$$Loss_{custom} = \lambda \times MSE(P_{pred}, P_{act}) + MSE(P_{pred}, P_{emp}) \quad (1)$$

where P_{pred} is the core loss predicted from the neural network, P_{emp} is the core loss from the empirical loss model (based on ISE with adaptive coefficients), and P_{act} is the measured core loss as given for each waveform in the training dataset.

The MSE indicates the mean-squared error; here, λ is again a hyperparameter set by employing the grid-search algorithm. This hyperparameter, λ , when fine-tuned, helps balance out the contribution from data loss and physics loss and aids in the proper training of the neural network.

The first term generally $MSE(P_{pred}, P_{act})$ corresponds to ‘*data-loss*’ since the output of the neural network is compared against the actual measured core loss values. The second term $MSE(P_{pred}, P_{emp})$ corresponds to ‘*physics-loss*’ or ‘*empirical-loss*’ since the output of the neural network is compared against the empirical equation. Hence, the idea of

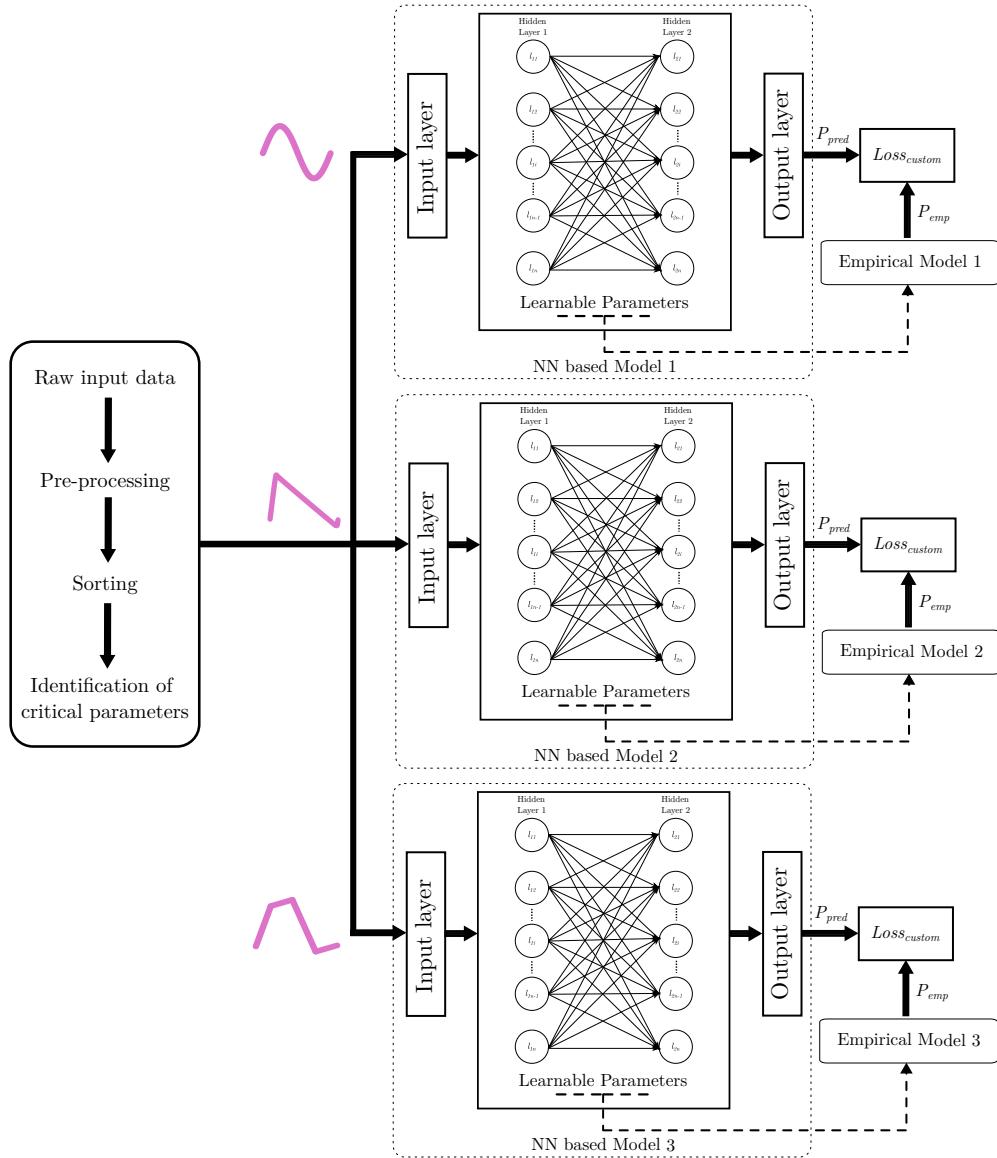


Fig. 2. Implementation of proposed hybrid model.

the hybrid model is inspired by PINN; empirical models are used instead of physics models.

Pre-training with pre-existing data of materials

The hybrid models have been pre-trained with the pre-existing data of 10 materials available on the MagNet Database. The pre-training of the neural network helps capture common core loss trends across the different materials. This boosts the learning of the neural network from the previous ten materials. It helps the models for the new material to achieve faster convergence and also prevents overfitting in the case of the materials where the number of data points in the training dataset is less.

Fine tuning and loss prediction

The pre-trained models and the learnable parameters are loaded into each neural network, thus initializing the weights

and biases to trained values rather than random values. The hyperparameters of the models are then fine-tuned for each material for the best training. A final model for each excitation type is saved for testing.

III. MODEL SUMMARY

Table II lists the architecture of the three models as shown in Fig. 2. The total number of parameters including the learnable parameters for each model is listed. These same models are used for different excitations of each of the 5 new materials.

Table III lists the 95th percentile error numbers for each excitation of the 5 new materials based on the training dataset provided. Table IV lists the final tuned hyperparameters for each of the models.

TABLE II
MODEL SIZE AND PARAMETERS FOR EACH EXCITATION AND MATERIALS

Model	Sinusoidal	Triangular	Trapezoidal
Size	One input layer: (7,16) Two hidden layers: (16,256), (256,16) One output layer: (16,1)	One input layer: (8,16) Two hidden layers: (16,256),(256,16) One output layer:(16,1)	One input layer: (12,16) Two hidden layers: (16,256), (256,16) One input layer: (16,1)
Parameters	8609	8625	8689

TABLE III

95TH PERCENTILE ERROR AFTER PRE-TRAINING AND FINE-TUNING OF THE MODELS WITH FULL 100% TRAINING DATASET

Material	Sinusoidal	Triangular	Trapezoidal
A	2.30 %	4.71 %	3.42 %
B	2.08 %	1.55 %	2.55 %
C	4.90 %	3.55 %	3.72 %
D	3.68 %	4.73 %	2.76 %
E	2.75 %	4.55 %	4.25 %

COMPEL 2020, ser. 2020 IEEE 21st Workshop on Control and Modeling for Power Electronics, COMPEL 2020. United States: Institute of Electrical and Electronics Engineers Inc., Nov. 2020.

- [7] T. Dimier and J. Biela, "Non-linear material model of ferrite to calculate core losses with full frequency and excitation scaling," *IEEE Transactions on Magnetics*, pp. 1–1, 2023.
- [8] T. Guillod, J. S. Lee, H. Li, S. Wang, M. Chen, and C. R. Sullivan, "Calculation of ferrite core losses with arbitrary waveforms using the composite waveform hypothesis," in *2023 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2023, pp. 1586–1593.

TABLE IV

THE FINAL TUNED HYPERPARAMETERS FOR EACH OF THE MODELS

Material	Models		
	Sinusoidal	Triangular	Trapezoidal
Hyperparameters			
A	LR_INIT = 5.6e-3 EPOCHS = 8000 BATCH SIZE = 128 λ = 215	LR_INIT = 5.6e-3 EPOCHS = 11000 BATCH SIZE = 128 λ = 215	LR_INIT = 5.6e-3 EPOCHS = 11000 BATCH SIZE = 128 λ = 215
B	LR_INIT = 5.5e-3 EPOCHS = 8000 BATCH SIZE = 128 λ = 460	LR_INIT = 5.5e-3 EPOCHS = 8000 BATCH SIZE = 128 λ = 460	LR_INIT = 5.6e-3 EPOCHS = 4000 BATCH SIZE = 128 λ = 460
C	LR_INIT = 6.5e-3 EPOCHS = 20000 BATCH SIZE = 256 λ = 80	LR_INIT = 6.0e-3 EPOCHS = 4000 BATCH SIZE = 128 λ = 132	LR_INIT = 5.5e-3 EPOCHS = 4000 BATCH SIZE = 128 λ = 132
D	LR_INIT = 3.0e-3 EPOCHS = 4000 BATCH SIZE = 64 λ = 193	LR_INIT = 3.0e-3 EPOCHS = 7000 BATCH SIZE = 64 λ = 193	LR_INIT = 3.0e-3 EPOCHS = 10000 BATCH SIZE = 64 λ = 193
E	LR_INIT = 5.5e-3 EPOCHS = 5000 BATCH SIZE = 128 λ = 1000	LR_INIT = 5.5e-3 EPOCHS = 7000 BATCH SIZE = 256 λ = 1000	LR_INIT = 5.5e-3 EPOCHS = 15000 BATCH SIZE = 256 λ = 1000

REFERENCES

- [1] D. Jiles and D. Atherton, "Theory of ferromagnetic hysteresis," *Journal of Magnetism and Magnetic Materials*, vol. 61, no. 1, pp. 48–60, 1986. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0304885386900661>
- [2] C. P. Steinmetz, "On the law of hysteresis," *Transactions of the American Institute of Electrical Engineers*, vol. IX, no. 1, pp. 1–64, 1892.
- [3] J. Reinert, A. Brockmeyer, and R. De Doncker, "Calculation of losses in ferro- and ferrimagnetic materials based on the modified steinmetz equation," *IEEE Transactions on Industry Applications*, vol. 37, no. 4, pp. 1055–1061, 2001.
- [4] K. Venkatachalam, C. 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., 2002, pp. 36–41.
- [5] S. Barg, K. Ammous, H. Mejbri, and A. Ammous, "An improved empirical formulation for magnetic core losses estimation under nonsinusoidal induction," *IEEE Transactions on Power Electronics*, vol. 32, no. 3, pp. 2146–2154, 2017.
- [6] H. Li, S. Lee, M. Luo, C. Sullivan, Y. Chen, and M. Chen, "Magnet: A machine learning framework for magnetic core loss modeling," in *2020 IEEE 21st Workshop on Control and Modeling for Power Electronics*,

Conditional Generative Adversarial Network Aided High-Frequency Core Loss Predictions

Xiaobing Shen, Qingcheng Sui, Jiaze Kong, and Wilmar Martinez

Abstract—This report addresses the intricate task of estimating magnetic iron losses in high-frequency (HF) magnetic components through an innovative approach, employing a newly developed Conditional Generative Adversarial Network (cGANET) model. Moving away from conventional loss estimation methods that often fail to consider the complex interactions among various factors, our cGANET model is meticulously engineered to integrate a wide array of elements such as material characteristics, shape discrepancies, and ambient factors. This advanced method is supported by the use of a specialized four-wire measurement apparatus, greatly enhancing the training dataset with its broad spectrum of measurements. In comparison to traditional Deep Neural Network (DNN) models, cGANET not only achieves quicker convergence but also excels in accuracy for predicting iron losses. This model's exceptional performance, especially in situations beyond the scope of the training data, highlights its robustness and flexibility. The improvements in predictive precision and operational efficiency mark a significant stride in the development and refinement of HF magnetic components.

Index Terms—High-Frequency Magnetic Cores, Volumetric Iron Losses, Conditional Generative Adversarial Network (cGANET), Deep Neural Network (DNN), Multi-layer perceptron (MLP)

I. INTRODUCTION

RECENT strides in modeling magnetic materials use analytical models to tackle multidimensional effects [1]–[3]. DNNs, Transfer Learning, and LSTM have been employed for data-driven research in HF magnetics [4]–[6].

Standard neural networks have limitations, particularly in predictions outside the training data range. To overcome this, we introduce a Generative Adversarial Network (GANET)-based model. GANETs have shown promise in various fields, including image generation and electromagnetics [7]–[10]. They can augment datasets, enhancing the robustness of predictive models.

Our GANET-based model aims to improve prediction accuracy, especially for operational points outside the training data.

This report contributes in the following ways:

- Prepares training data for the cGANET model.
- Validates a GANET-based model for training accuracy and extrapolation capabilities.
- Tests the model's generality on different materials and datasets.

II. PROPOSED GANET MODEL FOR IRON LOSS PREDICTIONS

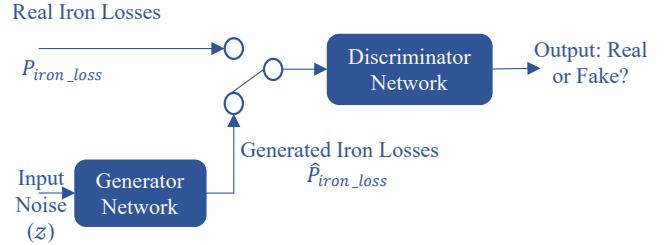


Fig. 1: Block Diagram of basic GANET

The GANET-based model surpasses the MLP model by enabling the creation of new data instances through innovative training techniques. Fig. 1 illustrates the GANET structure, consisting of the Generator, which synthesizes data akin to real instances, and the Discriminator, which discerns between real and synthetic data. The Generator's function is depicted as converting noise inputs (z) into data reflecting the desired distribution. The Discriminator's dual evaluation role is symbolized by a switch, underscoring its pivotal position in the GANET framework. The GANET training process is a strategic game where the Discriminator aims to accurately identify real data, counteracted by the Generator's attempts to produce increasingly convincing fakes. Fig. 2 introduces the Conditional GANET (cGANET), an evolution of the GANET model that incorporates a conditioning variable (x) into both Generator and Discriminator. Unlike standard GANETs that use single data points, cGANET employs pairs of input-output data ((x, y)), enriching the dataset and refining the training regime. The selection of cGANET for our predictive

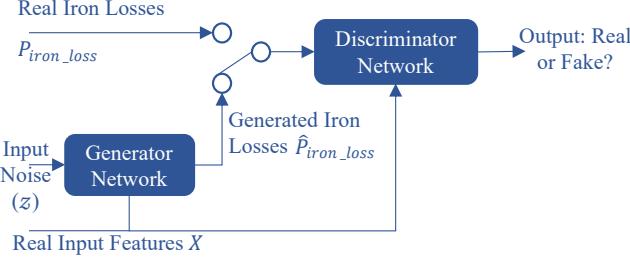


Fig. 2: Block Diagram of cGANET

modeling is due to its enhanced accuracy and context-awareness, essential for our application, as supported by [11].synthetic data.

A. Loss Functions of The Proposed GAN Model

A noise-injection architecture is used here. Within the framework of noise injection architecture, we introduce the noise denoted as z into the hidden layers at every level of the network. This mechanism is visually depicted in Fig. 3. It is worth noting that the dashed box, which signifies a concatenation block, is present in both Figure 3. The structure of the discriminator employed in our proposed model is depicted in Fig. 4. In a cGANET,

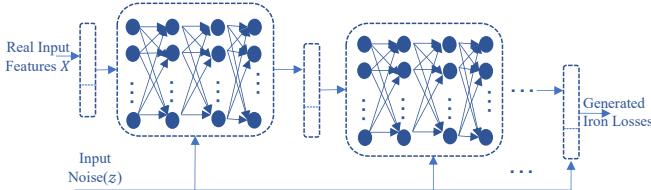


Fig. 3: Structure of The Noise-injection Generator

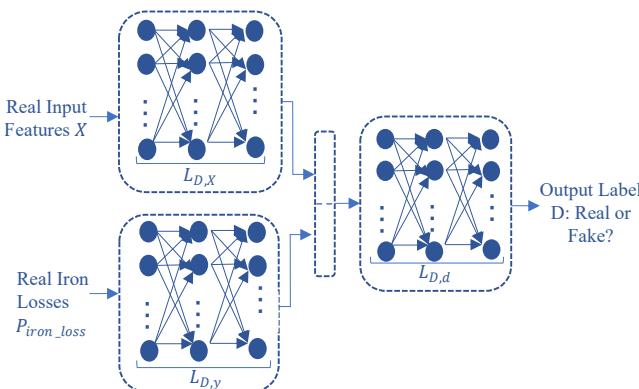


Fig. 4: Structure of The Noise-injection Discriminator

the generator and discriminator are conditioned on additional information such as class labels or data from other modalities. This additional information is provided

to both the generator and discriminator as input, which allows the cGANET to generate data that are conditioned on specific attributes.

The loss functions in a cGANET are modified to take into account the conditioning of the additional information. Here's a breakdown of the loss functions for both the discriminator and generator in a cGANET:

Discriminator Loss: The discriminator loss in a cGANET is similar to that of a traditional GANET, but with the condition information concatenated to the input. The loss is a sum of the losses for the real and generated (fake) data, and it's usually implemented as a binary cross-entropy loss. The discriminator wants to correctly classify real data as real and generated data as fake.

For a batch of data, the discriminator loss L_D can be formulated as [7]:

$$L_D = -\mathbb{E}_{\mathbf{x}, \mathbf{c} \sim p_{\text{data}}(\mathbf{x}, \mathbf{c})} [\log D(\mathbf{x}, \mathbf{c})] - \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z}), \mathbf{c} \sim p_{\text{data}}(\mathbf{c})} [\log(1 - D(G(\mathbf{z}, \mathbf{c}), \mathbf{c}))] \quad (1)$$

Where:

- D is the discriminator network.
- G is the generator network.
- \mathbf{x} is data sampled from the true data distribution $p_{\text{data}}(\mathbf{x})$.
- \mathbf{c} is the conditioning variable (like class labels).
- \mathbf{z} is the latent space vector sampled from distribution $p_{\mathbf{z}}(\mathbf{z})$ (typically a normal distribution).

Generator Loss: The generator's goal is to generate data that the discriminator will classify as real. In a cGANET, the generator is also conditioned on the additional information. The generator loss encourages the generator to create data that the discriminator will think is real [11].

The generator loss L_G can be expressed as:

$$L_G = -\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z}), \mathbf{c} \sim p_{\text{data}}(\mathbf{c})} [\log D(G(\mathbf{z}, \mathbf{c}), \mathbf{c})] \quad (2)$$

The generator tries to minimize this function. In practice, sometimes the generator loss is implemented as:

$$L'_G = \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z}), \mathbf{c} \sim p_{\text{data}}(\mathbf{c})} [\log(1 - D(G(\mathbf{z}, \mathbf{c}), \mathbf{c}))] \quad (3)$$

to provide stronger gradients during training, especially at the beginning when D is likely to reject samples from G with high confidence.

Overall Objective: The cGANET is trained by alternating between optimizing L_D and L_G (or L'_G). The complete min-max game between the generator and the discriminator can be summarized as follows:

$$\min_G \max_D L_{\text{cGANET}}(D, G) = L_D + L_G \quad (4)$$

(or $L_D + L'_G$ for the alternative generator loss).

This min-max game theoretically leads to a point where the generator produces perfect samples (indistinguishable from real data), and the discriminator is maximally confused, assigning a probability of 0.5 to both real and generated data. In practice, achieving this equilibrium is challenging and requires careful tuning of model architectures, learning rates, and other hyperparameters. At the beginning of the hyperparameters

TABLE I: Hyperparameters and Setups of cGANET-based Model

Hyperparameter	Value
Epochs	2000
Batch Size	32
Activation	ReLU
Performance	Relative Error
Learning Rate	Scheduler
Optimizer	Adam
Generator Parameters	118785 (464.00 KB)
Discriminator Parameters	99585 (389.00 KB)

tuning, some initial tests were applied to determine reasonable intervals for the hyperparameters. These initial setups consisted of shorter training runs with manually set parameters. The hyperparameters are set as in Table I. Then, the detailed structure of the hidden layers of the Generator and Discriminator were determined. The first two layers of the generator in Fig. 3 are with 128 nodes, connected to the following 4 layers with 64 nodes. For the Discriminator, $L_{(D,X)}$ and $L_{(D,y)}$ contain 2 layers with 64 nodes, followed by $L_{(D,d)}$ with 6 layers with 128 nodes in Fig. 4.

B. Results Analysis

Datasets from MagNet [12] are used to verify feasibility and generality, we use different training datasets from different materials. Four datasets, including Ferrites 3C90, 3C94, N87, and N49 from MagNet [12], were used to extend the evaluation of the proposed model. All the datasets were divided into two sets, one for the training process (75% of the whole set), and one beyond the former training set (the rest 25%) for the extra new test to verify its data augmentation and enhancement for predicting the scenarios beyond the training dataset.

For MagNet data, the input features include flux densities, frequencies, and temperature. The number of data items is smaller than that of our measurements, as shown in Table II. With limited data, the model still can work as a prediction model for totally new scenarios of one certain magnetic material. The training data range is 25 - 70°C, while the extra new test data is with 90°C, extending beyond the training set.

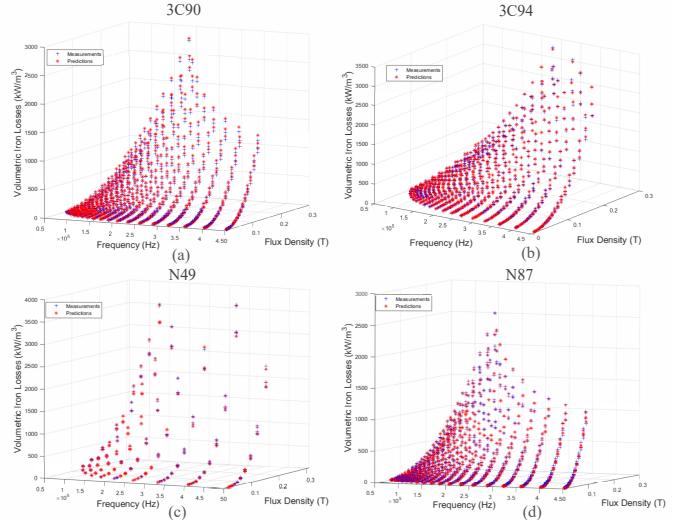


Fig. 5: Scatter Plot for Training Process of cGANET with: (a) 3C90, (b) 3C94, (c) N49, (d) N87 Datasets from MagNet

As the results of the cGANET model, Fig. 5 shows the scatter plot of predictions and measurements during the training process with these four datasets. N49 only has 249 data items for training, and it still can train the model well with an R-value of 0.9998. New datasets beyond the training data were applied after the proposed model was trained.

Fig. 6 shows scatter plots of predictions and measurements with new test data beyond the training set. All four materials can predict the unseen scenarios outside the range of the test data. All four plot results have R-values over 0.9900. The data augmentation and enhancement abilities are not affected by data sizes and sources. Table II presents the cGANET-based models in terms of training performance and new test experiments. This demonstrates cGANET's enhanced predictive ability in scenarios beyond the training range.

III. DATA PREPARATION FOR MAGNET CHALLENGE

The datasets encompass five provided materials, labeled as A, B, C, D, and E, each with a distinct count of data points in both training and testing phases.

The training dataset consists of five CSV files, featuring measurements for B_Field (in Teslas), H_Field (in Amps per meter), Frequency (in Hertz), Temperature (in degrees Celsius), and Volumetric_Loss (in Watts per cubic meter). The files follow a structure of $N \times 1024$ for B_Field and H_Field, and $N \times 1$ for Frequency, Temperature, and Volumetric_Loss, where N signifies the number of data points.

The distribution of data points for each material is as follows:

TABLE II: Overview of Datasets Used for Proposed cGANET-based Model

Materials	Source	Model	Data Items	R-value		Training Time (s)	Epoch
				Training	New Test		
3C90	MagNet	cGANET	1548	0.9997	0.9989	197.30	2000
3C94	MagNet	cGANET	1394	0.9998	0.9985	195.59	2000
N49	MagNet	cGANET	333	0.9998	0.9951	198.27	2000
N87	MagNet	cGANET	1603	0.9994	0.9972	198.55	2000

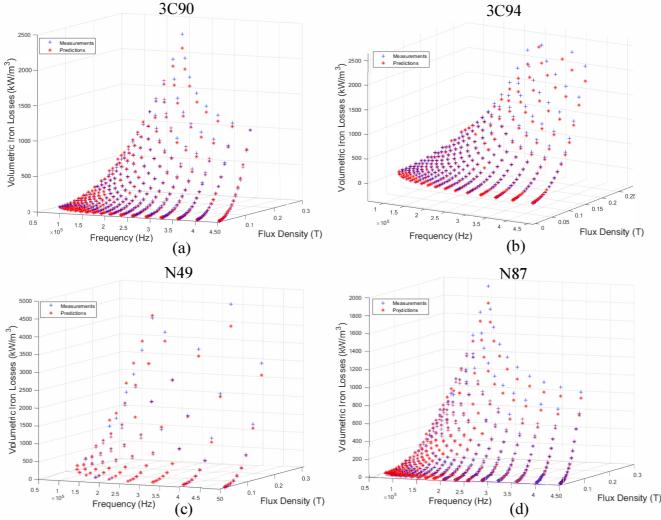


Fig. 6: Experimental Results of cGANET with New Data Tests of (a) 3C90, (b) 3C94, (c) N49, (d) N87 Datasets from MagNet

- Material A: 2432 data points for training, 7651 for testing.
- Material B: 7400 data points for training, 3172 for testing.
- Material C: 5357 data points for training, 5357 for testing.
- Material D: 580 data points for training, 7299 for testing.
- Material E: 2013 data points for training, 3738 for testing.

Our data preparation strategy can be conclude as follows:

- 1) **Waveform Categorization:** We classify different waveform shapes (sinusoidal, triangular, trapezoidal) as hot vectors, aiding in the analysis of waveform impact on magnetic properties.
- 2) **Feature Selection:** Our input features include flux density, frequency, and temperature, crucial in determining the magnetic behavior under different conditions.
- 3) **Target Value Identification:** The primary objective is to predict the volumetric core loss, our target

metric for assessing material efficiency.

- 4) **Data Normalization and Cleaning:** We normalize the data for comparability and clean it to address any inconsistencies or missing values.
- 5) **Feature Engineering:** We explore feature engineering techniques to potentially enhance model performance.

As part of our data preparation, we represent waveform types as hot vectors and include other parameters such as flux density, frequency, and temperature. Below is an example table showcasing the hot vector for waveforms. Our data preparation methodology is designed to

TABLE III: Example of Hot Vector Representation for Waveforms

Sinusoidal	Triangular	Trapezoidal
1	0	0
0	1	0
0	0	1

effectively handle the complexity and diversity of the provided datasets. By categorizing waveforms, selecting relevant features, and focusing on volumetric core loss prediction, we aim to develop robust models for accurate assessment of the magnetic properties of these materials under various conditions.

IV. RESULTS FOR MAGNET CHALLENGE DATA

After data preparation, the proposed cGANET model was trained with the provided five materials. The average relative error of the training process of each material can be found in Fig. 7.

The scatter plot with the R-value can be seen in Fig. 8. The results highly align with the results from Section II, which use the datasets from Magnet website with single excitation. Then our code is provided along with report, as well as the prediction for the test data.

The training process validates the use of cGANET for high-frequency magnetic core losses. And with the experiments in Section II confirm the prediction ability of the proposed cGANET model for the input beyond the training data range. Furthermore, with all the different materials and excitation waveforms, cGANET still holds

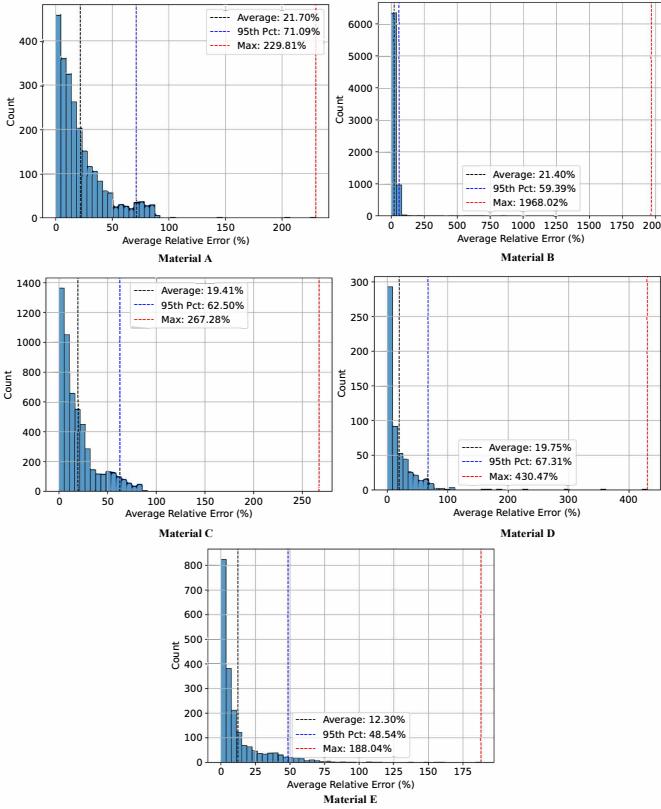


Fig. 7: Average Relative Errors of 5 Materials

the ability to provide accurate predictions. For Magnet Challenge, the flux density data doesn't have any noise ideally, which makes the training behavior a bit uglier, compared with the real data from the Magnet website. The concept of cGANETs applied to core loss prediction is appealing, providing accurate results on data for which the model has not been trained before. Besides, the impressive accuracy achieved even on sparse data, cGANETs effectively solve the problem associated with extrapolation on some conventional neural networks.

REFERENCES

- [1] N. Rasekh, J. Wang, and X. Yuan, "Artificial neural network aided loss maps for inductors and transformers," *IEEE Open Journal of Power Electronics*, vol. 3, pp. 886–898, 2022.
- [2] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen, and M. Chen, "Magnet: A machine learning framework for magnetic core loss modeling," in *2020 IEEE 21st Workshop on Control and Modeling for Power Electronics*, 2020, pp. 1–8.
- [3] E. Dogariu, H. Li, D. S. López, S. Wang, M. Luo, and M. Chen, "Transfer learning methods for magnetic core loss modeling," in *2021 IEEE 22nd Workshop on Control and Modelling of Power Electronics (COMPEL)*, pp. 1–6.
- [4] H. Li, D. Serrano, T. Guillod, E. Dogariu, A. Nadler, S. Wang, M. Luo, V. Bansal, Y. Chen, C. R. Sullivan, et al., "Magnet: An open-source database for data-driven magnetic core loss modeling," in *2022 IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 2022, pp. 588–595.
- [5] 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," *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14 292–14 316, 2023.
- [6] R. Y. M. Li, B. Tang, and K. W. Chau, "Sustainable construction safety knowledge sharing: A partial least square-structural equation modeling and a feedforward neural network approach," *Sustainability*, vol. 11, no. 20, p. 5831, 2019.
- [7] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sen-gupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, 2018.
- [8] Z. Qin, Z. Liu, P. Zhu, and W. Ling, "Style transfer in conditional gans for cross-modality synthesis of brain magnetic resonance images," *Computers in Biology and Medicine*, 2022.
- [9] O. Noakosteen, J. Vijayamohan, A. Gupta, and C. Christodoulou, "Antenna design using a gan-based synthetic data generation approach," *IEEE Open Journal of Antennas and Propagation*, vol. 3, pp. 488–494, 2022.
- [10] M. Baldan and P. Di Barba, "Discovering pareto-optimal magnetic-design solutions via a generative adversarial network," *IEEE Transactions on Magnetics*, vol. 58, no. 9, 2022.
- [11] J. Oskarsson, "Probabilistic regression using conditional generative adversarial networks," 2020.
- [12] Princeton-Dartmouth-Plexim MagNet Project, "Data-driven methods for power magnetics modeling," accessed: Jul. 8, 2023. [Online]. Available: <https://mag-net.princeton.edu/>

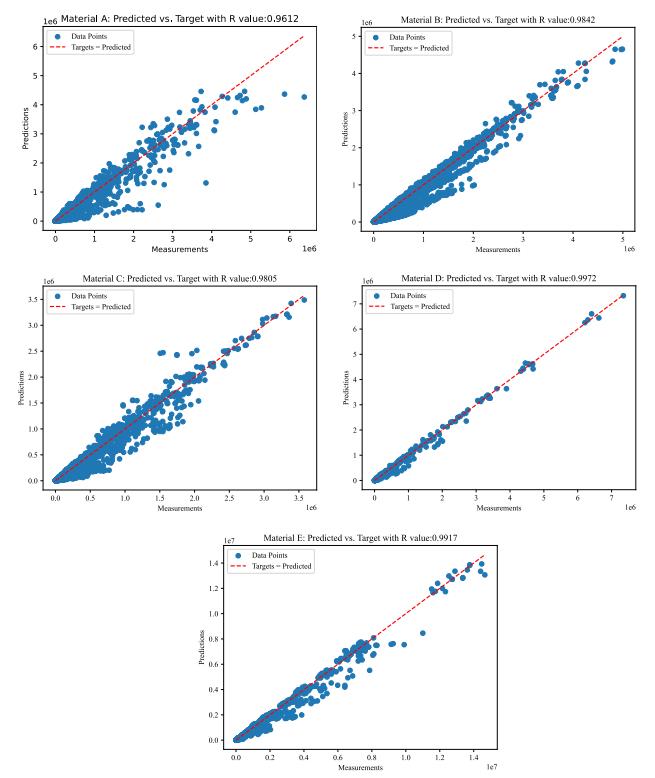


Fig. 8: Experimental Results of cGANET with 5 Materials from MagNet Challenge

Ferrite Core Loss Prediction with Direct Data Interpolation Method

Yidong Tian, *Student Member, IEEE*, and Cheng Zhang, *Member, IEEE*

Abstract- This paper proposes a direct data interpolation method for predicting ferrite core loss in power electronics. The core idea is to accurately construct the B-H loop shape given arbitrary H-field waveforms. Unlike conventional loss models like the Steinmetz equation and its variations, or recent neural network methods like MagNet that leverage periodic B-H waveforms and frequencies, this method employs a data segmentation and reconstruction process to fully exploit the distinctive features of the B-H loop under varying conditions. The source data is from Princeton University's MagNet open-source datasets.

I. INTRODUCTION

The magnetic components used in the power electronic systems are considered to have periodic flux waveforms in the core only. Current loss prediction methodologies are all based on this concept. In some variations, such as MSE/GSE, the flux waveform and its first-order derivatives are considered, based on the theory of the domain wall shifting [1, 2] that would well cover the hysteresis and eddy current losses. A third phenomenon of residual loss [2] has been picked up in i^2GSE [3]. In the development of these models, most formations, especially the first principles, are from observations of empirical relationships such as power law and exponential law, whilst some are from fundamental mathematical derivations such as harmonic analysis. However, the link between mathematical models and real-world materials is never accurate and reliable. At the time being, scientists are still not capable of predicting materials' physical properties on the macroscopic scale from knowing the formation of the matter. Not to mention the difficulties in controlling production. Practical materials, despite how much care has been provided, come with matter impurities and nonuniformities. These will add uncertainty to the applicability of traditional analytical formulae for core loss predictions.

The core loss prediction problem, from a generic engineering point of view, can be modeled as a canonical function depicted in Fig. 1, where two inputs, the time-varying waveform of \mathbf{B} , and the material's properties \mathbf{m} , are taken as the input and a single scalar value P_b , the volumetric loss power, is the output. Both \mathbf{B} and \mathbf{m} are not simple quantities but abstractive information packs, and in any implementation of this canonical function will have to discretize or summarize them. For instance, the original Steinmetz equation

$$P_\nu = k f^\alpha \hat{B}^\beta \quad (1)$$

is a simplification of

$$\begin{cases} \mathbf{B} = \{f, \hat{B}\} \\ \mathbf{m} = \{k, \alpha, \beta\} \end{cases} \quad (2)$$

With the advancement of more and more powerful and inexpensive computing devices, large-scale databases and data-driven predictions are both economically and timewise viable. One of the emerging methods is neural network (NN) and deep learning solvers such as MagNet [4-7]. An NN evaluator can also be modeled as such.

$$\begin{cases} \mathbf{B} = \{\{B_1, t_1\}, \{B_2, t_2\}, \dots \{B_n, t_n\}\} \\ \mathbf{m} = \{\text{pretrained NN}\} \end{cases} \quad (3)$$

The waveform is normally represented by discrete samples over a periodic cycle, whilst the material property is the pretrained NN – where the topology and selection of activate functions are manually designed, which is not straightforward, and the behavior may be unpredictable; and the individual cells' weight vectors are trained from a wide range of sampled data.

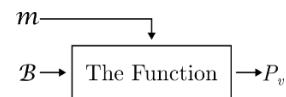


Fig. 1 The canonical function for the core loss prediction.

In [8], the authors asserted the two axioms for a plausible solution, which are adopted in this proposal:

Axiom 1: The loss must be a continuous function of parameters describing the waveform.

Axiom 2: If there are two equivalent descriptions of a waveform, calculations of loss based on either must give the same result.

Axiom 1 is difficult to strictly proven for a NN method, also can be seen from the unsmooth curves in [4]'s Fig. 17; but for axiom 2, any generic data presentation by sampling points should be naturally abiding.

This paper proposes a method that is similar to MagNet but can be formed more predictably and determined, whilst fully utilizing the sampled datasets. This results in a method that combines both the direct data interpolations and the database techniques. Data segmentation methods are used to determine the values in the database. In this paper, the methodology with detailed ideas and assumptions is described in Section II, Section III studies some cases in the Magnet datasets and shows the prediction results using the innovative data interpolation method. The conclusions are summarized in Section IV.

II. METHODOLOGY

A. Ideas and Assumptions

To accurately predict the ferrite core loss in power electronics, this paper implements a direct data interpolation method for constructing the related B-H loop shape with either arbitrary H-field or B-field waveforms. As mentioned in [4-6], the volumetric core losses per cycle can be calculated from the area enclosed by the B-H loop (energy per cycle) times the frequency, which is proportional to the integral of $H(t)$ over $B(t)$, or the integral of $B(t)$ over $H(t)$, as

$$P_v = f \oint H(B) dB = f \oint B(H) dH \quad (4)$$

For simplicity, in this paper, the arbitrary H-field waveforms are regarded as the inputs of the prediction method, since they are always plotted as the horizontal axis in the B-H loop figures.

In more detail, the direct interpolation is based on the hypothesis of dB/dH 's dependence on instant B , H , dH/dt , and temperature values, which is the theoretical basis for a series of hysteresis models named history-independent models [9, 10], such as the Duhem models [11]. Therefore, the predicted B-H loops can be constructed from accumulative extracted/interpolated dB/dH values in a database built from the training data segmentation. The growth of the B-H curve may start from an initial state, but eventually will be stabilized. It should be noted that, during the whole process, there are no extrapolation considerations, which means no “prediction” data is used for the B-H curve generation.

B. The Direct Data Interpolation Method

To better describe the developed direct data interpolation method, the following three propositions shall be clarified:

Proposition 1: For a labelled material, i. e. stable formula and production, the derivative of the B field with respect to the H field is dependent on the temperature, B field value, H field value, and the derivative of the H field with respect to time.

Therefore, the fundamental knowledge of core loss in a given material is a set of the following tuple mappings, as

$$(T, B, H, \frac{dH}{dt}) \rightarrow \left(\frac{dB}{dH} \right) \quad (5)$$

Proposition 2: Accurately measured B-H waveforms can be dismantled (the training data segmentation) to provide tuple mappings and saved in a big database (referred as bDB below). The coverage of the four-dimensional parameter space contributes to the accuracy of the prediction. The envelope should enclose the testing H field waveform to be simulated.

Proposition 3: Numerical integration using the interpolated dB/dH values with respect to the B, H, and dH/dt simulates the physical development of the B field from the H field.

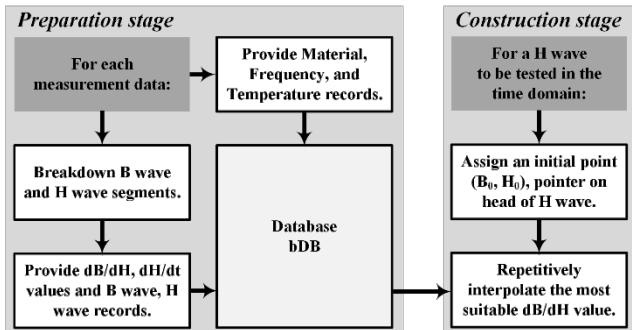


Fig. 2 The flow chart of the data interpolation method.

The flow chart of the data interpolation method is shown in Fig. 2. Overall, this method can be divided into two stages, which are the **Preparation** stage and the **Construction** stage. The steps in each stage are listed below:

Preparation:

Step 0: Inputs. Material, B wave, H wave, frequency, and temperature records.

Step 1: For each record, breakdown B and H segments, and compute dB/dH and dH/dt values, and store into bDB.

Construction:

Step 0: Inputs. H waveform to be tested in the time domain, repetitive indefinitely.

Step 1: Assign an initial point (B_0, H_0) , pointer on head of H waveform.

Step 2: Interpolated dB/dH value w.r.t. current $(T, B, H, dH/dt)$ from bDB.

Step 3: Move to the next step $(B, H, dH/dt)$.

Step 4: Go to *Step 2*, or stop after multiple complete cycles when the B-H loop is stable.

The most basic assumption in this method is that the B-H hysteresis can be piecewise constructed. Therefore, dB/dH is the critical data to be predicted. It is unclear whether it is only dependent on the T , B , H , or dH/dt values, or any combinations of them, but there is no harm to use them all. The four independent variables in equation (5) are called “coordinates”, and dB/dH is the “altitudes” to be interpolated from the four dimensional “map”. Because all data points are from existing measures, the behaviour should be identical to the real material. This can also be considered as a giant look-up table (LUT).

However, the data points in most situations are neither evenly distributed, nor evenly covered along all axes. There might be some areas with extreme dense points and some else very sparse. The challenge of this method is really how to effectively extract a number of most close data points near the given coordinates, which will be discussed in the next section with the MagNet dataset analysis and time complexity.

C. Data Analysis

The source data used in this paper is from Princeton University's MagNet open-source datasets. The parameters for each data record are shown in Table I.

TABLE I
THE PARAMETERS FOR EACH DATA RECORD IN MAGNET [5]

Parameter	Data Type	Description
Material	String	Name of material
Temperature	Real	in degree Celsius
Frequency	Real	in Hz
H-Field waveform	Real[1024]	in A/m
B-Field waveform	Real[1024]	in T
P_v	Real	in W/m ³

It is worth noting that the MagNet dataset is enormous. For each of the materials, there are some 40k data records, at four

temperatures, having a wide range of frequencies from 49.95kHz to 794.34kHz and various excitation current waveforms. Each record contains 1024 points, therefore, at least 1023 segments can be safely used, and that results in about 40M individual data points for one material (or about 10M points for a measured temperature). Finding the nearest n points normally requires to compute the Euclidean distance between the point and all points in the bDB. Even disregard sorting (that gives n nearest points), going through the whole bDB alone takes O(n) time, which is unacceptable.

Therefore, a method that effectively selects a range of nearest points within O(log n) time is proposed. In *Step 1* in the **Preparation** stage when data points are inserted into bDB, an index is inserted at the same time - a composite index of (T, B, H, and dH/dt), in a tree structure (B-Tree) where sorted and ranged queries can be executed in O(log n) time. A modern SQL database is suitable for such purpose. Also, a dynamic range with auto sizing (feedback control) is used to limit the sizes of points to be pulled out from bDB.

D. Loss Analysis

The data acquisition system for MagNet datasets is illustrated in [5]’s Fig. 3 and Fig. 4, in which the testing setup uses a gapless circular coil with two windings. The temperature of the core is controlled with water/oil bath thermostat, and the frequency is controlled by the function generator. Then, when the core is a simple shape like a doughnut with a uniform cross-sectional area of A and an average circumference length of l, and N turns of the winding, the amplitude of the H-field wave can be projected from the current measured through the first winding (excitation), as

$$H = \frac{N}{l} I_1 \quad (6)$$

The amplitude of the B-field wave can be calculated from the integration of the voltage measured from the secondary winding (capture), as

$$B = -\frac{1}{NA} \int V_2 dt \quad (7)$$

Therefore, it can be considered that the enclosed area by B-H loop is essentially the $I_1 \cdot V_2 dt$ loop. Assume the I_2 (current in the secondary loop) is almost 0, this loss is recognized as equivalent to the total energy lost in the magnetic core, including the hysteresis loss, eddy current loss, and other losses.

III. CASE STUDY

The measurement data of N87 material in Magnet is studied in this section, which can fully demonstrate the capabilities of the proposed direct data interpolation method for both ferrite core loss prediction and B-H curve construction. For N87 material, there are 40616 training data and 5000 validation data in total. As shown in Table I, during each B-H curve construction, there are at least 1024 interpolation steps to guarantee a complete B-H loop. Then, since in (4), the core loss is proportional to the B-H loop area, the relative error of each time core loss prediction can be defined as

$$E_{cost} = \frac{Area_{BHpred} - Area_{BH}}{Area_{BH}} \times 100\% \quad (8)$$

where $Area_{BHpred}$ is the area of the method constructed B-H loop, and $Area_{BH}$ is the area of the measured B-H loop provided by MagNet. Also, refer to the practical situation, the B-H loop construction are all start from the (0, 0) point.

A. Training Data B-H Loop Recovery

As mentioned in Section II.B, in the method **Preparation** stage, a database bDB is built to save all the provided information of the N87 training data, as well as their dB/dH and dH/dt values. Therefore, the direct data interpolation method is first performed on the training data itself, to demonstrate the B-H loop construction effect when having all the curve details.

Three B-H loop recovery results are shown in Fig. 3, with completely different temperature, frequency, and B-H loop shape parameters. It can be seen that, when the B-H loop details exist in the database, through the (T, B, H, dH/dt) coordinates, the interpolation method can well pinpoint the most suitable dB/dH value at each construction step, therefore the B-H loops can be completely recovered.

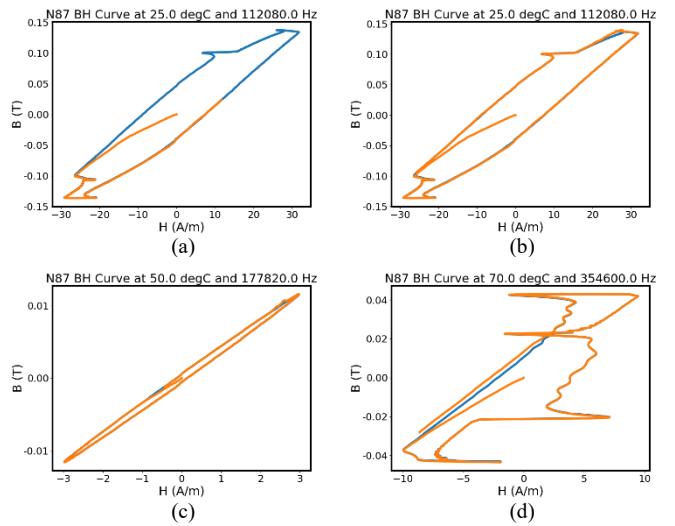


Fig. 3 The recovery results for three different B-H loops belonging to the N87 training data (blue curve represents the measured B-H loop and orange curve represents the constructed B-H loop).

B. Data Validation

Then, the data interpolation method is executed on the 5000 N87 validation data provided by MagNet. The relative error of the predicted core loss is calculated using equation (8). And the error analysis relative to temperature, frequency, and core loss are shown in Fig. 4, respectively.

For time and accuracy considerations, only 4847 out of 5000 data are considered in the core loss error analysis. All of them have finished at least 1100 steps (97% finished over 2100 steps) in three minutes (executed on a laptop having Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz, 16 GB RAM), and have a relative core loss error below 95%. The average core loss error is 19.42%. As shown in Fig. 4, the errors are uniformly distributed over temperature and frequency, but have clear distribution characteristics relative to core losses: the validation data with lower core losses are more likely to have larger core loss errors.

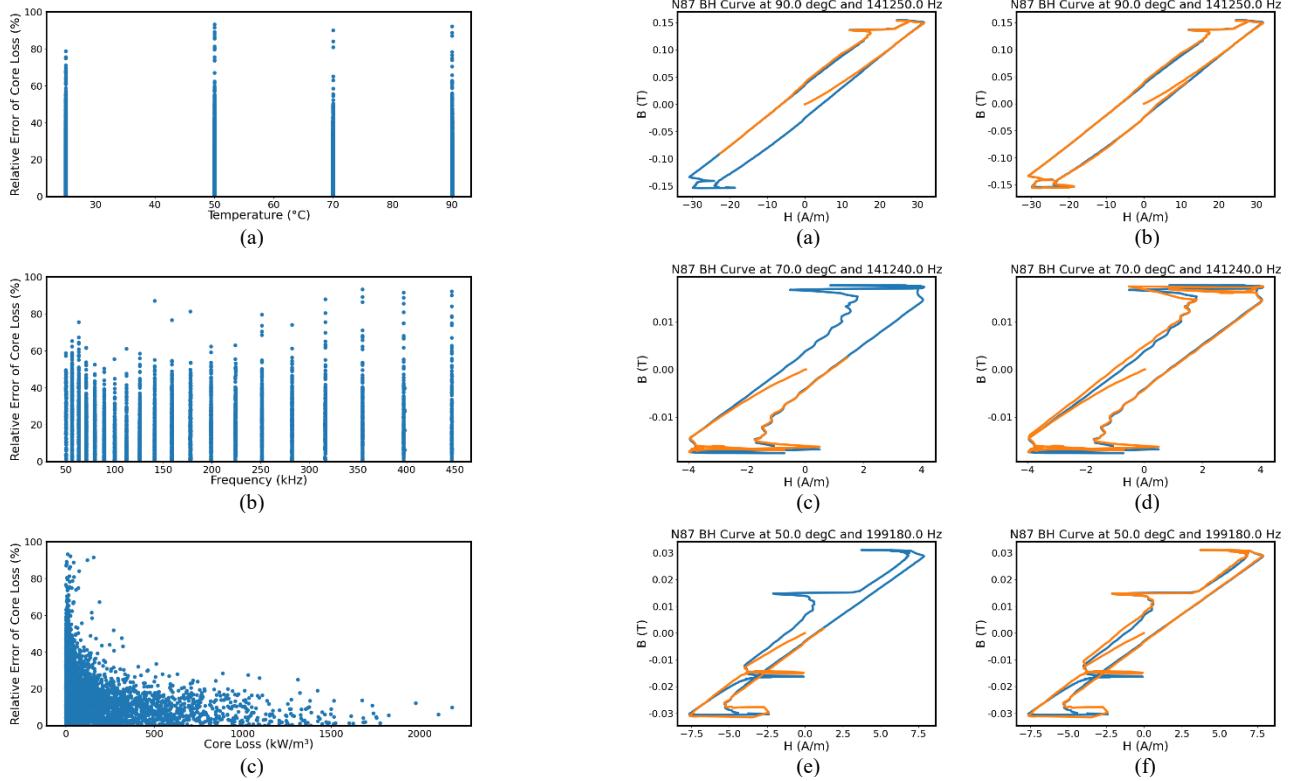


Fig. 4 The relative core loss error analysis relative to (a). temperature, (b). frequency, and (c). core loss.

In more detail, four B-H loop construction results are shown in Fig. 5, with different relative core loss errors. From the smallest to the largest, Fig. 5 (a) (b), (c) (d), (e) (f), (g) (h) show the B-H loops with predicted core loss errors of 1.92%, 12.31%, 24.61%, and 44.83%, respectively. It can be seen that, the core loss error mainly comes from the misalignment of certain envelope segments between the constructed B-H curve and the measured one. Meanwhile, when the B-H loop area is small (the core loss is small) but the envelope misalignment is obvious, the error can become quite large. This is consistent with the error analysis results in Fig. 4.

C. Discussion

During the B-H loop construction, there are two phenomena that need to be discussed. The first phenomenon is shown in Fig. 6, where the constructed curve collapses at the B-H loop's upper edge. This is because at this edge, the theoretical dB/dH value shall be 0. However, in the bDB database, the dB/dH values at this edge which are calculated from the measured training data can have both negative and positive values. Therefore, if there is a large H segment, the B value can then have an obvious change. This discussion is also suitable for the B-H loop's lower edge.

The second phenomenon is shown in Fig. 7, where the constructed curve fails to track the envelope of the B-H loop. This is because there is no required data in the bDB database. As mentioned in Section II.A, in this method, no “prediction” data is used for B-H loop construction.

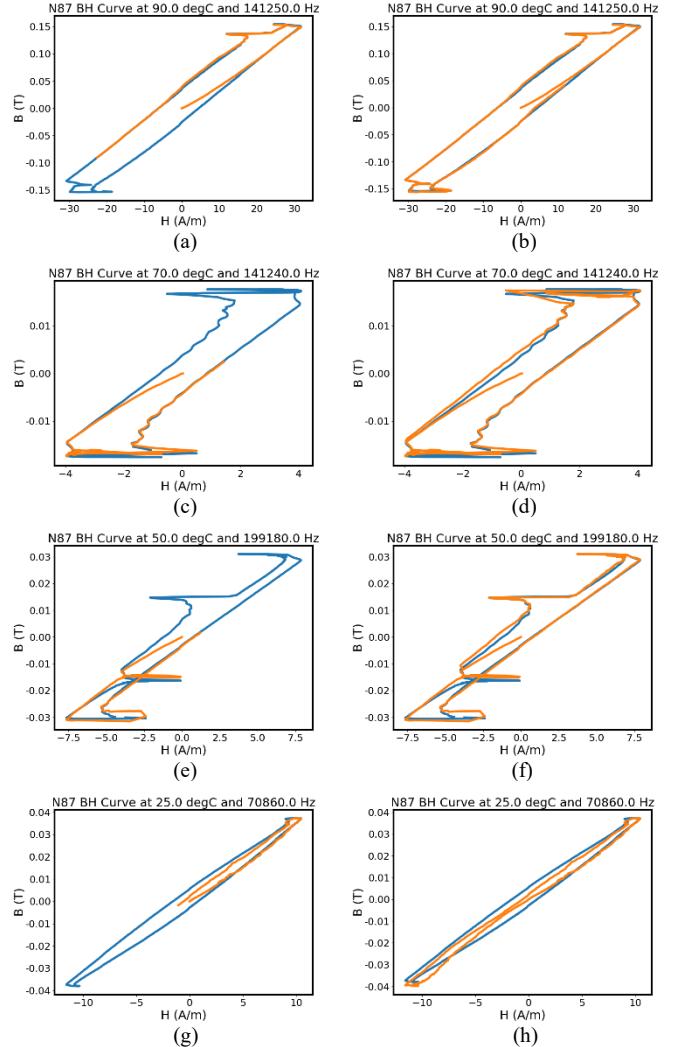


Fig. 5 Four B-H loop construction results with different relative core loss errors. (a) and (b). 1.92%, (c) and (d). 12.31%, (e) and (f). 24.61%, (g) and (h). 44.83%.

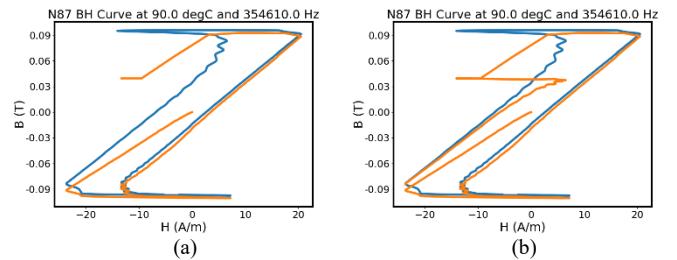


Fig. 6 The first phenomenon: the constructed curve collapses at the B-H loop's upper edge.

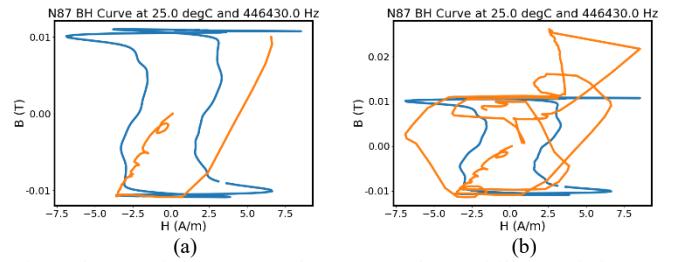


Fig. 7 The second phenomenon: the constructed curve fails to track the envelope of the B-H loop.

IV. CONCLUSION

A direct data interpolation method is proposed in this paper to predict the ferrite core loss in power electronics, which is based on the accurate construction of the B-H loop shape given arbitrary H-field waveform. Through using Princeton University's open-source MagNet datasets, the data interpolation method is both verified on the N87 material training datasets and validation datasets, and shows good ferrite core loss prediction results under different frequency, temperature, and B-H loop shape conditions.

REFERENCES

- [1] 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, July-Aug. 2001, doi: 10.1109/28.936396.
- [2] J. B. Goodenough, "Summary of losses in magnetic materials," in IEEE Transactions on Magnetics, vol. 38, no. 5, pp. 3398-3408, Sept. 2002, doi: 10.1109/TMAG.2002.802741.
- [3] 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, doi: 10.1109/TPEL.2011.2162252.
- [4] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen and M. Chen, "MagNet: A Machine Learning Framework for Magnetic Core Loss Modeling," 2020 IEEE 21st Workshop on Control and Modeling for Power Electronics (COMPEL), Aalborg, Denmark, 2020, pp. 1-8, doi: 10.1109/COMPEL49091.2020.9265869.
- [5] H. Li et al., "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," in IEEE Transactions on Power Electronics, vol. 38, no. 12, pp. 15829-15853, Dec. 2023, doi: 10.1109/TPEL.2023.3309232.
- [6] D. Serrano et al., "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, doi: 10.1109/TPEL.2023.3291084.
- [7] 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, doi: 10.1109/TPEL.2023.3309233.
- [8] K. Venkatachalam, C. R. Sullivan, T. Abdallah and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," 2002 IEEE Workshop on Computers in Power Electronics, 2002. Proceedings., Mayaguez, PR, USA, 2002, pp. 36-41, doi: 10.1109/CIEP.2002.1196712.
- [9] S. E. Zirka, Y. I. Moroz, P. Marketos and A. J. Moses, "Congruency-based hysteresis models for transient simulation," in IEEE Transactions on Magnetics, vol. 40, no. 2, pp. 390-399, Mar. 2004, doi: 10.1109/TMAG.2004.824137.
- [10] N. Vaiana, R. Capuano, and L. Rosati, 2023. "Evaluation of path-dependent work and internal energy change for hysteretic mechanical systems." In Mechanical Systems and Signal Processing, vol. 186, pp.109862, 2023, doi: <https://doi.org/10.1016/j.ymssp.2022.109862>.
- [11] U. Hornung, "The mathematics of hysteresis," in Bulletin of the Australian Mathematical Society, vol. 30, no. 2, pp. 271-287, Oct. 1984, doi: <https://doi.org/10.1017/S0004972700001957>.

The composite improved improved Generalized Steinmetz Equation (ci²GSE): Pushing Analytical Models to the Limit

Asier Arruti , Borja Alberdi, Anartz Agote, Iosu Aizpuru 

Abstract- This document presents a compact core loss model designed to be presented in magnetic material datasheets built upon well known works regarding the topic, while adhering to the MagNet challenge evaluation criteria. The inputs have been modified to better fit a typical magnetic device design workflow, where the flux waveforms are typically simplified versions. Then the developed model is presented, which is built upon the basis of loss separation into hysteresis and eddy losses, composite waveform hypothesis, and relaxation losses. All these effects are modelled in a small number of parameters, 3 for the hysteresis losses, 3 for the eddy losses and 3 for the relaxation losses. Lastly the obtained results are presented, followed by a discussion on why we believe this model is an excellent candidate for magnetic material datasheets.

Index Terms- Core loss modelling, MagNet challenge.

I. INTRODUCTION

The following document represents the MagNet challenge final report presented by the Mondragon University team.

Our objective in the MagNet challenge was to generate a model that is simple enough to be presented in magnetic material datasheets while following the MagNet challenge evaluation criteria:

- Model accuracy (30%): although in the MagNet challenge the 95th percentile is the criteria to evaluate, we decided to focus in minimizing the root mean square error instead, since we believe that the 95th percentile goes against the concept of generalizable models, which should be the key criteria in the generation of a standard material loss definition.
- Model size (30%): the model should be as compact as possible, making a significant trade of accuracy vs size if this means that the model can be defined in a small number of parameters ideal for publishing in the material datasheets along loss curves.
- Model explainability (20%): the model is based on well known analytical approaches for loss calculation (hysteresis loss, eddy loss and relaxation loss), [1]-[4], facilitating its adoption in industry and combination with other works based on the Steinmetz Equations (like the Steinmetz Premagnetization Graph [5]).
- Model novelty (10%): model size and accuracy have the same weight in the evaluation criteria (30% and 30%).

Asier Arruti, Borja Alberdi, Anar Agote and Iosu Aizpuru are with the Faculty of Engineering, Mondragon University, Mondragon, Spain (e-mail: arrutti@mondragon.edu; janzola@mondragon.edu; iaizpuru@mondragon.edu; mmazuela@mondragon.edu).

but we believe that most teams will focus on the accuracy part, while our approach will focus in generating a compact and generalizable model that can be easily presented in material datasheets.

– Software quality (10%): the model should be easy to understand and implement in different environments (even in basic tools such as EXCEL).

As the title suggest, the proposed model is based on analytical models, mainly using 4 main pillars:

- I. True Steinmetz Equation (tSE): assumption that two terms are necessary to accurately define the core loss behavior at low (hysteresis losses) and high (eddy losses) frequencies [1].
- II. The improved generalized Steinmetz Equation (iGSE): assumption that the losses of non-sinusoidal waveforms can be defined using the flux derivative (dB/dt) and peak to peak flux density (ΔB) of the waveform [2].
- III. The composite waveform hypothesis (CWH): assumption that in triangular/trapezoidal flux waveforms the losses of each segment can be calculated separately [3].
- IV. The improved improved generalized Steinmetz Equation (i²GSE): assumption that trapezoidal waveform with zero voltage segments increase the total losses due to relaxation effect [4].

The composite improved improved Generalized Steinmetz Equation (ci²GSE) presented in this work is the natural continuation of the composite improved Generalized Steinmetz Equation (ciGSE) model first conceptualized in [6]. The ciGSE was created as an accurate alternative to the iGSE for high and low duty cycle triangular waveforms, combining the CWH and iGSE in a way that allows to retain the direct connection with the Steinmetz Parameters. The preliminary report sent on the 10th of November details the limitations of the ciGSE found during the MagNet challenge, which have been completely solved in this iteration of the model:

- I. The characterization of the loss model is fully automated and 100% generalizable between materials and temperatures. The automatic parametrization can predict the optimal parameters from the same initial points for all materials.
- II. The extrapolation problem has been completely solved with the use of the dual loss plane based on the tSE, where two planes are used to define the low frequency (hysteresis) and high frequency (eddy) losses. This also

reduces the amount of parameters necessary substantially, but entails a loss in accuracy in the low to high frequency transition. Still, the final accuracies are still impressive for the very compact model achieved, which is in accordance with our defined criteria.

The following report will describe the presented model in various sections. First, in Section II the necessary changes of the input data to use the model will be described. Then, in Section III the main model is presented, with the necessary steps to parametrize the model and representation of the different physical phenomena governing core losses. Then, on Section IV the results achieved with the model for the preliminary 10 materials and the final 5 materials are presented. Lastly, a discussion about the model is presented in Section V followed by the conclusions in Section VI.

II. INPUT DATA STRUCTURE

Since the ciGSE was initially developed to be used in the design process of transformers for core loss estimations, the inputs disagree with the ones presented by the MagNet challenge and resemble more the simplified waveforms segments typically used in classical approaches such as the iGSE. The MagNet challenge present experimental flux waveforms, while in the design process these are not available.

Thus, the flux waveforms must be transformed into basic segments, requiring a preprocessing of the data. The algorithm proposed to do so first classifies the flux waveforms into three categories:

- I. Sinusoidal waveforms: the fast Fourier transform is applied to the flux waveform, and if the main harmonic is much higher than the rest the waveform is classified as sinusoidal.
- II. Triangular waveforms: the derivatives of the flux waveforms are obtained and if they have only two levels, the waveform is classified as triangular. The duty cycles are obtained from the derivative of the flux density.
- III. Trapezoidal waveforms: same approach as triangular waveforms but if three levels are detected the waveform is trapezoidal. The duty cycles are also obtained from the derivative of the flux density.

Note that this classification is not perfect, and sometimes can result in wrong classification of the waveform. Also, the duty cycles are defined in a resolution of 0.1, which appears to be the resolution used in the MagNet database.

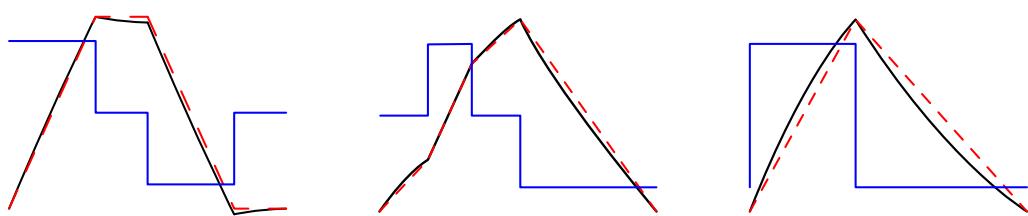


Fig. 1: Examples of two trapezoidal and triangular waveforms transformed in their equivalent simplified waveforms. Original flux waveform (**black**), flux derivative (**blue**) and reconstructed simplified waveform (**red**).

After the classification, simplified waveforms based on segments are generated to use with the proposed model. Examples of these simplified waveforms are shown in Fig. 1, a trapezoidal waveform with relaxation, a trapezoidal waveform without relaxation, and a triangular waveform. It is clear that these waveforms are not exactly the same as the non simplified flux waveforms, but are more akin to typical waveforms found in the design process of magnetic devices.

The necessary parameters to define these simplified waveforms then become:

Sinusoidal	Frequency	[Nx1]
	Peak to Peak Flux	[Nx1]
Triangular	Frequency	[Nx1]
	Duty cycle	[Nx2]
	Flux change	[Nx2]
	Flux derivative	[Nx2]
Trapezoidal	Frequency	[Nx1]
	Duty cycle	[Nx4]
	Flux change	[Nx4]
	Flux derivative	[Nx4]

where N is the number of waveforms.

REMARK: This initial preprocessing algorithm is a very simple approach, and will work in most cases, but from the initial 10 materials it has been seen that it sometimes fails in the classification of 1% to 2% of the waveforms. This would not be too much of a problem if the models are optimized for the 95th percentile, but since we decided to use the root mean square error, this data classification problem has an impact in the overall results. It should also be noted that at high frequencies the transition of the flux density derivative is not instant and using the 0.1 duty cycle resolution results in the simplified waveform not completely resembling the original waveform. Unfortunately, optimizing this algorithm would take too much time and effort away from the development of the main model, since the only way to correctly verify this would be to one by one plot the original and simplified waveforms and check for discrepancies.

III. THE ci²GSE MODEL

As mentioned in the introduction, the proposed ci²GSE is based on 4 critical pillars from existing literature of analytical modelling of core losses.

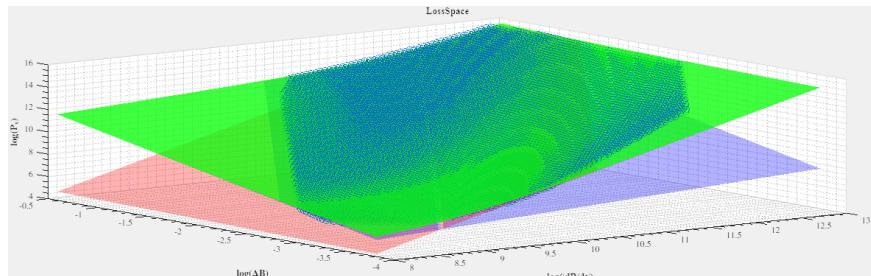


Fig. 2: Losses of each waveform segments as functions of flux density derivative and peak to peak flux density. Here the low frequency hysteresis losses (blue plane) and high frequency eddy losses (red plane) can be clearly visualized.

Beginning from the first one, the true Steinmetz Equation (tSE) which is the original Steinmetz equation presented in [5] to model the core losses, looks like this:

$$P_{\text{loss}} = k_1 f B^{b_1} + k_2 f^2 B^2 \quad (1)$$

Here, the first part refers to the hysteresis losses and the second term to the high frequency eddy losses. This can be rewritten in a similar manner to the iGSE, so it is defined as functions of the flux derivative dB/dt and peak to peak flux density ΔB :

$$P_{\text{loss}} = \frac{1}{T} \int_0^T \left[k_1 \left| \frac{dB}{dt} \right|^{a_1} \Delta B^{b_1} + k_2 \left| \frac{dB}{dt} \right|^{a_2} \Delta B^{b_2} \right] dt \quad (2)$$

The introduction of the two adding terms in the core loss definition is critical to represent the influence of the non-constant Steinmetz Parameters, meaning that unlike in the iGSE, if we have two segments with very different dB/dt values, the impact of the low frequency hysteresis losses and high frequency eddy losses are modeled independently in each segment. This is how the findings from the CWH [3] can be considered in our model, allowing to estimate the losses of high and low duty cycle waveforms much more accurately than the classical iGSE approach. This same concept was reported in our previous publication of the ciGSE [6].

Now, in our previous work [6] we used fifth degree polynomial surfaces (where $X Y Z$ are the logarithmic values of dB/dt , ΔB and P_{loss} respectively) to model the losses in each segment, but although this results in very accurate estimations, it generates problems when extrapolation is used. The new proposed model (2) is less accurate but can define the losses with only 6 parameters and should mathematically be able to semi-accurately predict the losses at ultra-low and ultra-high frequencies as shown in our preliminary report.

The task in hand becomes in how to use the MagNet data to obtain the necessary parameters k_1 , a_1 , b_1 , k_2 , a_2 and b_2 . What we propose in our model is to use all triangular data and trapezoidal data without relaxation effect to fit these parameters minimizing the total root mean square error. To do so we found that rewritten (2) into the following form helps the fitting algorithm find the optimal parameters,

$$P_{\text{loss}} = \sum D \left[\exp \left(k'_1 + a_1 \ln \left| \frac{dB}{dt} \right| + b_1 \ln \Delta B \right) + \exp \left(k'_2 + a_2 \ln \left| \frac{dB}{dt} \right| + b_2 \ln \Delta B \right) \right] \quad (3)$$

where D is the duty cycle, and the summation term represented the addition of the losses of all segments. Note that the ΔB term is not the flux density change in a segment, but the peak to peak flux density in the waveform, as defined according to the iGSE in [2].

The idea behind this concept is that the core losses are generally quite linear in logarithmic dimensions, thus using k' , a and b the loss space is a combination of two planes. This is better visualized in Fig. 2 where the core losses (per segment) in logarithmic dimensions are shown to fit two planes. The blue plane represents the low frequency losses (hysteresis losses) while the red plane represents the high frequency losses (eddy losses). The green plane is the combination of both planes, fitting the MagNet data quite closely.

From physical understanding of the hysteresis and eddy losses, we can make some approximations of the initial k'_1 , a_1 , b_1 , k'_2 , a_2 and b_2 parameters. We know that for the low frequency losses (blue plane) ideally the a_1 parameters should be close to 1 since the hysteresis losses are assumed to increase proportionally with frequency. For the high frequency losses, we have more information, since we know that the eddy losses are a function of the square of dB/dt , thus a_2 should be close to 2 while b_2 should be approximately 0. We can also make estimations of the value of k'_2 , since this value should relate to the resistivity of the core material, which for ferrites is commonly around 1 to 10 Ωm , meaning k_2 values of 1e^{-6} or 1e^{-7} , or k'_2 values around -13.8 and -16.1.

Based on our analysis of the initial 10 materials, we proposed using initial values of $k'_1 = 5$, $a_1 = 0.75$, $b_1 = 1.75$, $k'_2 = -15$, $a_2 = 2$ and $b_2 = 0$. Then, we can let the computer try to minimize the resulting root mean square error from (3) to find the ‘optimal’ values. Different algorithms can be used for this optimization, although we found that the basic MATLAB’s function to find the minimum of an unconstrained multivariable function (`fminunc`) works correctly.

Of course, due to the relaxation effect reported in [4], this approach will not be able to accurately estimate the losses in trapezoidal waveforms like the first one shown on Fig. 1. To solve this issue, a new term must be added to (3) to consider the increased losses. To do so, we first need to quantify the discrepancy in losses due to relaxation effect, which we can do by comparing the estimated losses using (3) and the real

240 measured losses from the MagNet data. From here we can
 250 define the relaxation energy:

$$E_{rel} = \frac{(P_{real} - P_{pred})}{f} \quad (4)$$

251 where P_{real} are the losses given in the MagNet data and P_{pred}
 252 are the losses estimated using (3). Once the relaxation energies
 253 E_{rel} are obtained, we can try to generate a mathematical
 254 definition for these. Although this problem was already tackled
 255 in the i²GSE [4], the equation presented there is not that easy to
 256 implement for thousands of waveforms at different frequencies
 257 and flux density levels. Based on the assumption that the
 258 relaxation losses are governed by mainly two parameters, the
 259 flux density ΔB and the relaxation time t_{rel} , we found that

$$E_{rel} = \exp(k'_{rel} + a_{rel} \ln t_{rel} + b_{rel} \ln \Delta B) \quad (5)$$

260 is in most cases good enough to achieve decent predictions of
 261 the increase in core losses. In this case, since (5) also represents
 262 a plane in logarithmic space similar to those shown in Fig. 2,
 263 the parameters can be extracted easily without requiring to
 264 define the initial points. Still, logic dictates that, according to
 265 the findings from [4], the a_{rel} parameters should be between 0
 266 and 1 since the relaxation energy increases with the relaxation
 267 time (>0) but in a decaying exponential function (<1).

268 With (3) and (5) we have everything needed to evaluate all
 269 trapezoidal and triangular waveforms, but since we have
 270 obtained the necessary parameters in two steps, we run into the
 271 risk of not finding the globally optimal parameters. Thus, as a
 272 last step the parameters should be reoptimized using a
 273 combination of (3) and (5):

$$P_{loss} = \sum \left[D \left(\exp \left(k'_1 + a_1 \ln \left| \frac{dB}{dt} \right| + b_1 \ln \Delta B \right) \right. \right. \\ \left. \left. + \exp \left(k'_2 + a_2 \ln \left| \frac{dB}{dt} \right| + b_2 \ln \Delta B \right) \right) \right] \\ + f \exp(k'_{rel} + a_{rel} \ln t_{rel} + b_{rel} \ln \Delta B) \quad (6)$$

274 where we can use the already obtained parameters as initial
 275 points.

276 According to our findings this last optimization is capable
 277 to reduce the root mean square error by up to x0.70 in some
 278 cases. The 95th percentile error usually also decreases by around
 279 x0.95 in most cases, but in some situations this last optimization
 280 step can increase the 95th percentile by up to x1.05 since we still
 281 optimize the parameters for the minimum root mean square
 282 error.

283 Note that until now only triangular and trapezoidal
 284 waveforms have been analyzed. For sinusoidal waveforms we
 285 opted to simply fit the sinusoidal loss data into a second-degree
 286 polynomial plane where X Y Z are the logarithmic values of f ,
 287 ΔB and P_{loss} respectively. This polynomial plane is then
 288 defined with 6 parameters: $p00, p10, p01, p20, p11$ and $p02$.

289 In the future we hope to use the well established physical
 290 connection between the iGSE and classical sinusoidal loss
 291 Steinmetz Equation to use the same parameters for sinusoidal,
 292 triangular, and trapezoidal waveforms.

IV. OBTAINED RESULTS

293 With the model presented, TABLE I presents the fitting
 294 parameters and results achieved for the preliminary 10
 295 materials.

TABLE I: RESULTS FOR THE PRELIMINARY 10 MATERIALS

Mat	Temp	k'_1 $p00$	a_1 $p10$	b_1 $p01$	k'_2 $p20$	a_2 $p11$	b_2 $p02$	k'_{rel} 60 parameters, 4x(9+6)	a_{rel}	b_{rel}	$E_{95\text{th}}$
3C90	25°C	3.2949	0.9926	1.5387	-18.0861	2.5519	-0.1489	9.4304	0.5558	2.2764	
	50°C	1.9388	1.1001	1.6632	-18.4833	2.5520	-0.3359	10.8116	0.6566	2.5594	
	70°C	1.8146	1.1170	1.9345	-18.8307	2.5006	-0.3363	11.6266	0.7022	2.8748	
	90°C	0.5324	1.2379	2.0636	-17.8128	2.4918	-0.3789	14.3733	0.9090	3.0760	
	All	3.1540	1.0761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	15.87%
	25°C	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	
3C94	50°C	1.7237	1.0830	1.6383	-15.7103	2.3611	-0.1608	9.2033	0.5471	2.5406	
	70°C	0.7116	1.1702	1.9055	-14.9323	2.2931	-0.1706	8.6424	0.4904	2.9140	
	90°C	-1.7024	1.3837	1.9884	-15.0473	2.2875	-0.2278	9.7426	0.5575	3.2584	
	All	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	15.87%
	25°C	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	
	25°C	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	
3E6	50°C	1.4756	1.0670	1.4486	-9.0378	1.9104	0.0629	8.3426	0.5284	1.7438	
	70°C	0.9241	1.1102	1.2908	-8.4753	1.8720	0.0795	8.1050	0.5144	1.6260	
	90°C	-0.3152	1.2309	1.0344	-9.6984	1.9477	-0.0126	7.4695	0.4876	1.4387	
	All	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	13.27%
	25°C	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	
	25°C	3.1540	0.9761	1.4614	-16.0243	2.4074	-0.0485	8.8663	0.5214	2.2247	
3F4	50°C	4.6360	1.0032	1.9359	-21.6308	2.7726	-0.1841	10.3548	0.5705	2.2826	
	70°C	4.1859	1.0438	1.9569	-24.7371	2.9376	-0.5927	11.1837	0.6283	2.3416	
	90°C	4.0261	1.0642	1.9695	-24.2859	2.8819	-0.2945	12.0015	0.6747	2.4572	
	All	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	6.048%
	25°C	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	
	25°C	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	
77	50°C	1.3287	1.1311	1.4277	-17.3259	2.4909	-0.2305	9.4489	0.5778	2.3135	
	70°C	-0.0006	1.2417	1.6690	-16.7898	2.4360	-0.2945	11.2179	0.7019	2.7183	
	90°C	-2.3497	1.4469	1.6627	-17.0067	2.4349	-0.3837	8.8988	0.5056	2.9511	
	All	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	14.34%
	25°C	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	
	25°C	3.7371	0.9372	1.4237	-16.1308	2.4407	-0.0044	8.5796	0.5016	2.1180	
N27	50°C	1.2854	1.1428	1.5804	-17.1200	2.4738	-0.2441	10.5646	0.6514	2.4860	
	70°C	-0.8585	1.3263	1.7031	-17.2370	2.4621	-0.3446	12.1486	0.7657	2.8140	
	90°C	-2.1911	1.4453	1.7032	-17.2121	2.4516	-0.3835	8.1462	0.4389	2.9331	
	All	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	14.34%
	25°C	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	
	25°C	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	
N30	50°C	4.3365	0.9017	1.4604	-15.3586	2.3917	0.0633	8.5848	0.4827	2.1194	
	70°C	1.9238	1.0985	1.5119	-16.6799	2.4481	-0.1812	9.6089	0.5707	2.3300	
	90°C	1.0170	1.1764	1.7118	-16.2404	2.4023	-0.2360	10.5126	0.6317	2.6017	
	All	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	14.48%
	25°C	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	
	25°C	3.0371	1.0024	1.4597	-16.8690	2.4838	-0.0916	9.4090	0.5610	2.2070	
N49	50°C	2.3451	1.1476	2.1030	-25.5739	2.9597	-0.6906	10.4937	0.5804	2.6461	
	70°C	2.2009	1.1701	2.1078	-25.8101	2.9847	-0.7025	10.4108	0.5649	2.6640	
	90°C	2.2715	1.1731	1.9736	-25.4128	2.9808	-0.6315	9.8899	0.5170	2.6323	
	All	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	9.774%
	25°C	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	
	25°C	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	
N87	50°C	2.7582	1.0433	1.5589	-18.0318	2.5658	-0.2361	10.9395	0.6584	2.3829	
	70°C	1.8971	1.1157	1.6869	-17.7556	2.5284	-0.3041	11.9197	0.7292	2.5825	
	90°C	0.8429	1.2119	1.8316	-17.2386	2.4832	-0.3470	13.4851	0.8385	2.8382	
	All	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	9.397%
	25°C	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	
	25°C	2.9209	1.0927	1.9977	-23.9579	2.8626	-0.5196	10.2641	0.5681	2.5259	

TABLE II: FITTING RESULTS FOR THE FINAL 5 MATERIALS

Mat	Temp	k'_1	a_1	b_1	k'_2	a_2	b_2	k'_{rel}	a_{rel}	b_{rel}	E_{95th}
Material E			p_{00}	p_{10}	p_{01}	p_{20}	p_{11}	p_{02}			
	25°C	2.3859	1.1314	1.9734	-29.1130	3.2238	-0.5807	11.6155	0.6969	2.2377	
			19.6599	-1.8043	2.6016	0.1299	0.0197	-0.1340			
	50°C	1.4351	1.2128	2.1435	-32.3441	3.4352	-0.7411	7.7280	0.3912	2.4369	
			18.0486	-1.4739	3.2517	0.1125	-0.0639	-0.3650			
	70°C	1.7758	1.1882	2.1923	-32.5128	3.4874	-0.4860	11.4230	0.6770	2.3816	
			14.3141	-0.8558	3.6948	0.0829	-0.1775	-0.9002			
	90°C	1.9566	1.1807	2.1885	-33.1304	3.5205	-0.5941	11.2878	0.6594	2.3600	
	All				36.9005	-5.1469	2.2420	0.2855	-0.0357	-0.1442	
			60 parameters, 4x(9+6)						23.60%		
Material D			p_{00}	p_{10}	p_{01}	p_{20}	p_{11}	p_{02}			
	25°C	3.5849	1.0122	1.7976	-19.0609	2.5856	-0.7373	0	0	0	
			11.3000	-0.4019	3.0320	0.0670	-0.0746	-0.1627			
	50°C	1.2973	1.1999	1.8173	-27.8189	3.1403	-0.8648	13.3134	0.9304	1.0925	
			15.3094	-1.6607	0.9518	0.1511	0.2150	0.1075			
	70°C	2.0268	1.1467	2.0771	-27.8331	3.1534	-0.7978	-0.2020	-0.3647	3.3655	
			-12.6105	4.3570	14.2491	-0.1631	-1.0063	-0.5620			
	90°C	-1.9950	1.4827	2.1022	-28.7020	3.1720	-0.9693	56.7664	3.7447	5.2287	
	All			31.3326	-3.5993	6.4942	0.2038	-0.2248	0.0109		
			60 parameters, 4x(9+6)						22.87%		
Material C			p_{00}	p_{10}	p_{01}	p_{20}	p_{11}	p_{02}			
	25°C	-0.8090	1.2789	1.3626	-18.9273	2.6114	-0.3846	5.5581	0.3012	2.5923	
			26.7373	-3.3736	4.1984	0.2043	-0.1506	-0.0248			
	50°C	-0.7740	1.2812	1.5520	-18.1526	2.5238	-0.4213	6.4205	0.3511	2.7855	
			41.6026	-6.0729	4.7102	0.3290	-0.1567	0.0543			
	70°C	-0.4455	1.2615	1.8683	-18.0382	2.5584	-0.3343	5.5969	0.2720	3.0418	
			31.1284	-4.0782	7.5609	0.2357	-0.3864	0.0504			
	90°C	-1.8676	1.3913	1.8703	-18.4802	2.5937	-0.3934	9.0339	14.0935	4.8501	
	All			31.4340	-4.2160	7.2207	0.2471	-0.3361	0.1190		
			60 parameters, 4x(9+6)						14.60%		
Material B			p_{00}	p_{10}	p_{01}	p_{20}	p_{11}	p_{02}			
	25°C	-0.6082	1.2897	1.8224	-12.5537	2.1550	-0.1191	5.6830	0.2541	2.9877	
			4.5861	0.3641	5.0111	0.0509	-0.1938	0.0717			
	50°C	-0.0972	1.2595	1.8690	-11.8026	2.1140	-0.0925	3.6906	0.0925	3.0152	
			0.9186	1.0051	4.8621	0.0243	-0.1831	0.0804			
	70°C	1.1251	1.1786	2.0075	-10.7112	2.0481	-0.0328	4.3712	0.1330	2.8837	
			-2.6573	1.7074	5.0152	-0.0087	-0.2046	0.0653			
	90°C	-0.1711	1.2885	1.7722	-9.7093	1.9818	-0.0017	4.3702	0.1647	2.5736	
	All			-.9.0400	2.8436	4.9328	-0.0588	-0.2100	0.0434		
			60 parameters, 4x(9+6)						8.074%		
Material A			p_{00}	p_{10}	p_{01}	p_{20}	p_{11}	p_{02}			
	25°C	4.7442	0.8548	1.4052	-18.7122	2.7643	-0.2571	2.4296	0.0808	2.4503	
			6.9005	-5.1469	2.2420	0.2855	-0.0357	-0.1442			
	50°C	3.1228	0.9707	1.3666	-19.1510	2.8185	-0.3063	5.0173	0.2651	2.8145	
			45.3397	-6.7173	3.1549	0.3577	-0.1022	-0.1063			
	70°C	-2.4755	1.4555	1.0968	-21.0851	2.9037	-0.5311	5.9455	162.8406	36.2542	
			36.9005	-5.1469	2.2420	0.2855	-0.0357	-0.1442			
	90°C	-6.8894	1.8204	1.2179	-18.4851	2.7187	-0.3902	-3.2823	106.0626	24.0666	
	All			35.5932	-5.4185	3.7661	0.3210	-0.0771	0.0755		
			60 parameters, 4x(9+6)						18.24%		

For the final 5 materials, we cannot present the 95th percentiles of the validation data, but since in the original 10 materials we have seen that the errors in the model fitting dataset and the model validation dataset are very similar, we can present the fitting errors instead. Similar to TABLE I, the fitting results for the final 5 materials are shown in TABLE II.

REMARK: For the final 5 materials the code for the original 10 materials has been repurposed and executed directly without any modifications. It is very possible that the erroneous data classification from the preprocessing stage behaves differently since the final 5 datasets are composed of very different waveform type distributions, thus being a potential major source of the increased 95th percentile error compared to the original 10 materials. The relation between the 95th percentile and root mean square error appears to indicate that this is the case, since in the original 10 materials the 95th percentile is almost twice of the root mean square error, while in the final 5 materials it is up to thrice the root mean square error.

V. DISCUSSION

The model that we have presented here offers a very compact analytical approach to model core losses with close ties to the known physical phenomena governing the hysteresis, eddy, and relaxation losses. We have proven that the model is 100% generalizable between materials since it does not require any kind of fine tuning or redefinition of the loss equations depending on the material and temperature.

Since the model is built upon well-known analytical approaches [1]-[4], we hope that other works built upon the same basis, such as the Steinmetz Premagnetization Graph presented in [5], can be intuitively integrated in the presented approach to model the behavior of other phenomena like premagnetization losses.

It is important to mention that in this work we took the approach of generating new parameters for each temperature, but looking at the parameters from TABLE I, there appear to be clear relations between the evolution of the parameters with the temperature. Looking at the k'_1 , it appears that it almost always decreases with temperature, meaning that the hysteresis loss plane is lowered at high temperatures. This would make sense with the physical behavior of the magnetic domains inside the material, since at higher temperatures the easier it would be for these domains to move and rotate, generating lower hysteresis losses. Thus, there is reason to believe that only one set of these material parameters could be defined for a given temperature, and that the parameters at different temperatures could potentially be obtained from modelling this behavior with the temperature. This could potentially reduce the number of parameters even further. Integrating the sinusoidal losses should not be too hard either with the well-known relation between the sinusoidal and triangular losses, thus we could reduce the number of required parameters even further.

To finish, we believe that the model proposed here would be ideal to be used as an standard kind of datasheet for material losses; it is very compact and should easily fit in a traditional datasheet, it is built upon well known literature, is tied with physical phenomena, and is really easy to implement in any software.

VI. CONCLUSION

This work represents the final report from the Mondragon university team for the MagNet challenge. The criterion behind the key decision to generate the model are justified with the aim of creating a standard approach to parametrize magnetic materials.

First, a modification of the input variables is presented, where the waveforms are transformed into simplified versions, which would be beneficial for using the model in a classical transformer design workflow. Then, the basis of the model are presented, which is built upon well known analytical approaches, combining the key concept of hysteresis and eddy loss separation, definition of losses using flux derivative and peak to peak flux density, utilization of the composite waveform hypothesis and lastly integration of the relaxation losses.

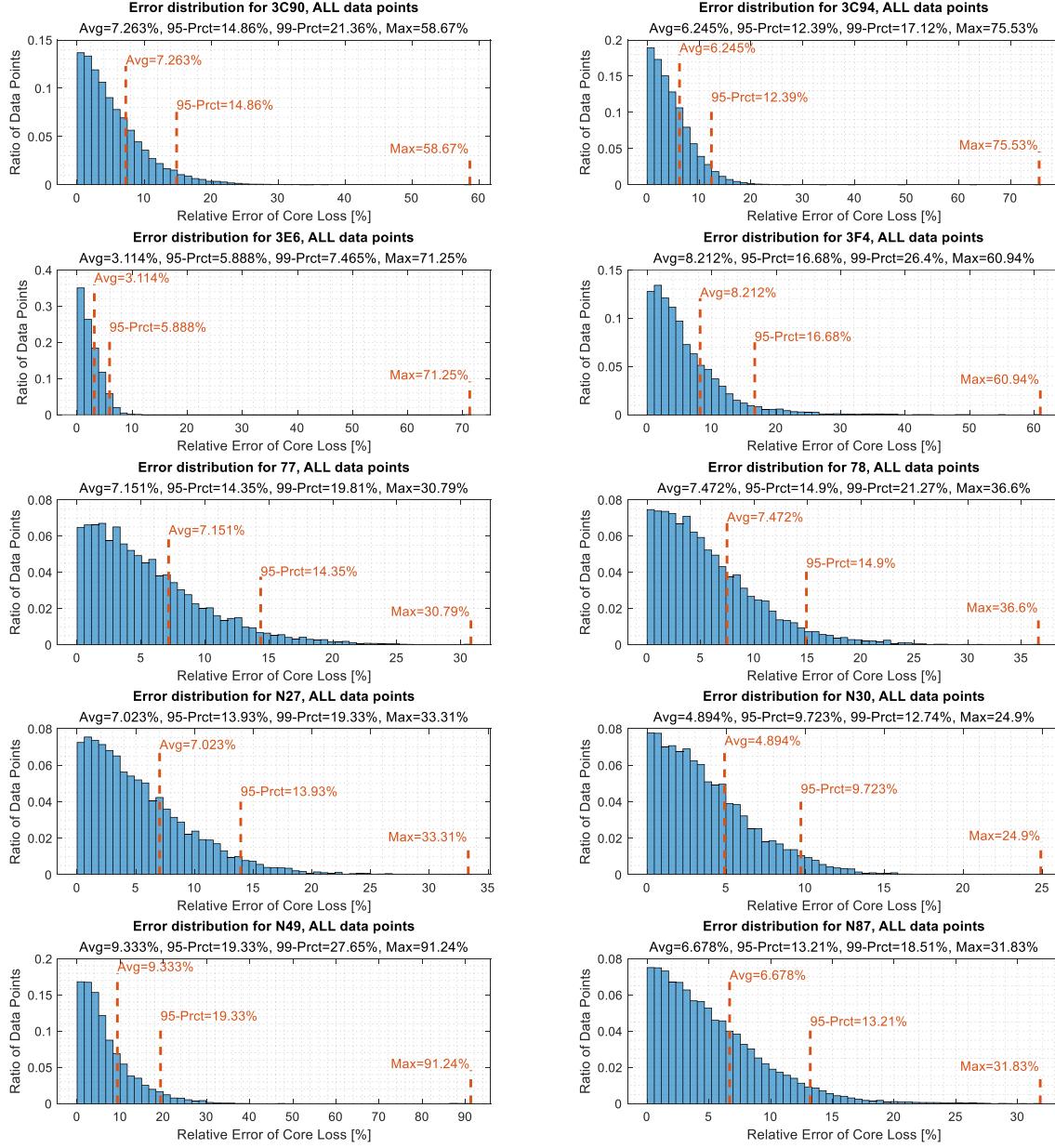
After this, the results obtained are presented, showing all the parameters used, as well as the 95th percentile errors achieved. Estimation of the results for the final 5 materials are also presented.

Lastly, a discussion is presented on why we believe that the presented model is a good candidate for material datasheets.

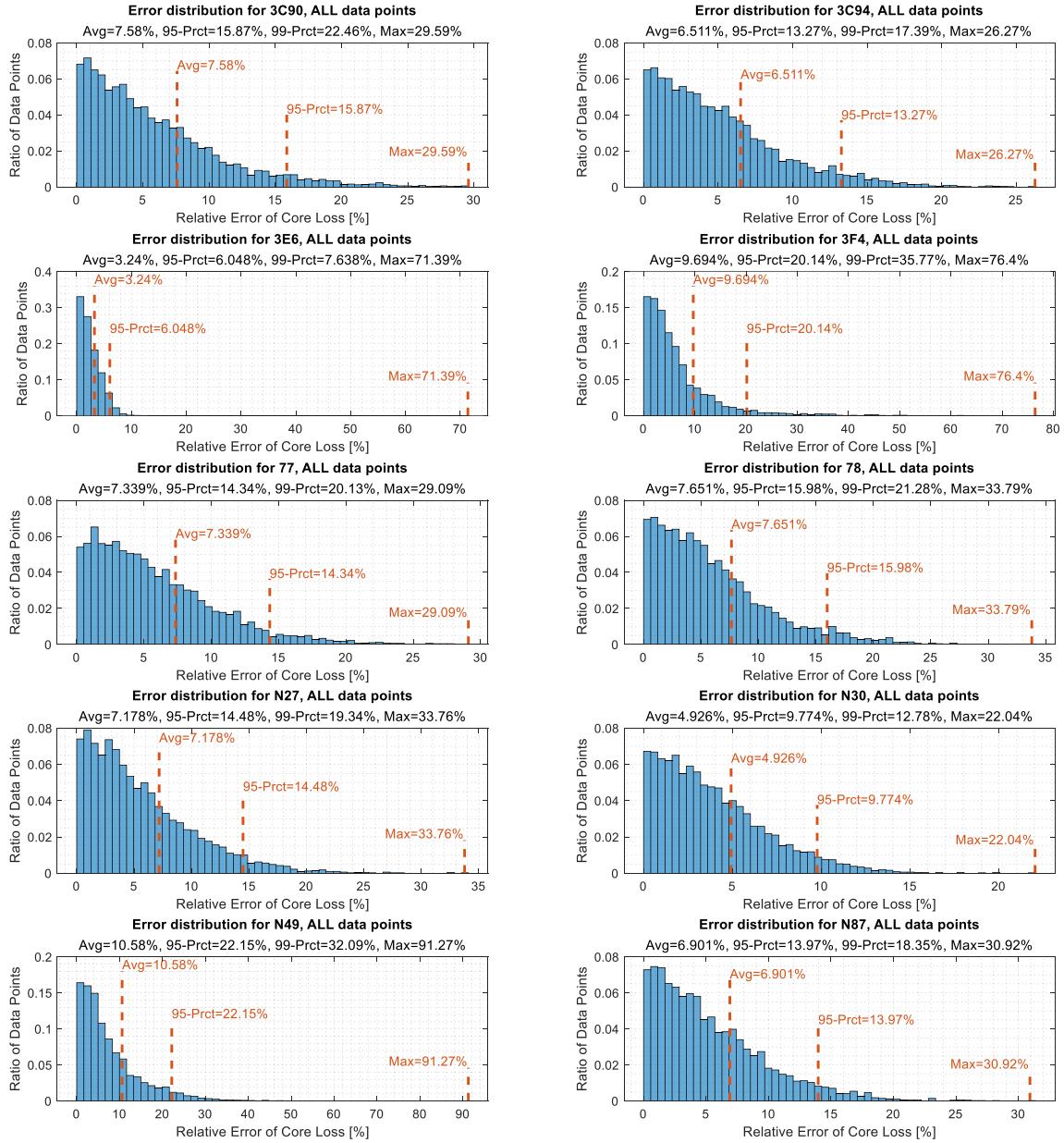
REFERENCES

- 377 [1] C. P. Steinmetz, "On the Law of Hysteresis," in *Transactions of the*
378 *American Institute of Electrical Engineers*, vol. IX, no. 1, pp. 1-64, Jan. 1892,
379 DOI: [10.1109/T-AIEE.1892.5570437](https://doi.org/10.1109/T-AIEE.1892.5570437).
- 380 [2] K. Venkatachalam, C. R. Sullivan, T. Abdallah and H. Tacca,
381 "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using
382 only Steinmetz parameters," 2002 IEEE Workshop on Computers in Power
383 Electronics, 2002. Proceedings., Mayaguez, PR, USA, 2002, pp. 36-41, DOI:
384 [10.1109/CIPE.2002.1196712](https://doi.org/10.1109/CIPE.2002.1196712).
- 385 [3] C. R. Sullivan, J. H. Harris and E. Herbert, "Core loss predictions for
386 general PWM waveforms from a simplified set of measured data," 2010
387 Twenty-Fifth Annual IEEE Applied Power Electronics Conference and
388 Exposition (APEC), Palm Springs, CA, USA, 2010, pp. 1048-1055, DOI:
389 [10.1109/APEC.2010.5433375](https://doi.org/10.1109/APEC.2010.5433375).
- 390 [4] J. Muhlethaler, J. Biela, J. W. Kolar and A. Ecklebe, "Improved Core-
391 Loss Calculation for Magnetic Components Employed in Power Electronic
392 Systems," in *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 964-
393 973, Feb. 2012, DOI: [10.1109/TPEL.2011.2162252](https://doi.org/10.1109/TPEL.2011.2162252).
- 394 [5] J. Muhlethaler, J. Biela, J. W. Kolar and A. Ecklebe, "Core losses
395 under DC bias condition based on Steinmetz parameters," *The 2010*
396 *International Power Electronics Conference - ECCE ASIA* -, Sapporo, Japan,
397 2010, pp. 2430-2437, DOI: [10.1109/IPEC.2010.5542385](https://doi.org/10.1109/IPEC.2010.5542385).
- 398 [6] A. Arruti, J. Anzola, F. J. Pérez-Cebolla, I. Aizpuru and M. Mazuela,
399 "The Composite Improved Generalized Steinmetz Equation (ciGSE): An
400 Accurate Model Combining the Composite Waveform Hypothesis With
401 Classical Approaches," in *IEEE Transactions on Power Electronics*, vol. 39,
402 no. 1, pp. 1162-1173, Jan. 2024, DOI: [10.1109/TPEL.2023.3323577](https://doi.org/10.1109/TPEL.2023.3323577).

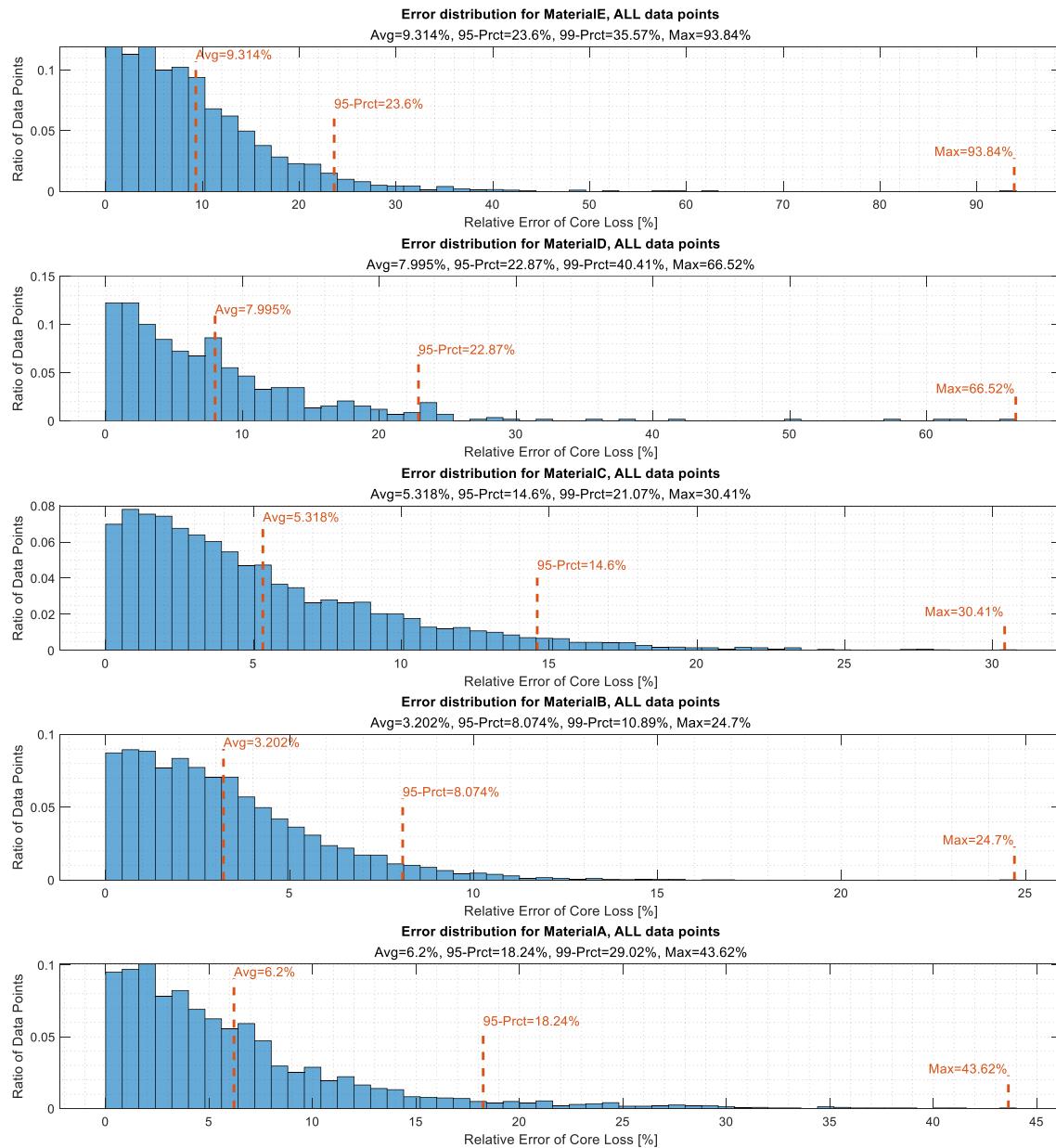
DETAILED RESULTS FOR THE INITIAL 10 MATERIAL FITTING



DETAILED RESULTS FOR THE INITIAL 10 MATERIAL VALIDATION



DETAILED RESULTS FOR THE FINAL 5 MATERIAL FITTING



Predicting Core Losses based on Data-driven Physical Model

(Concept Novelty Track for 2023 MagNet Challenge)

Zhang Minmin, Xia Chen, Dong Yuchen, Wang Duo, Shen Tianming, Zhou Yan*

College of Automation, Nanjing University of Posts and Telecommunications, Nanjing
210046, China

Abstract- Core losses are tightly related with the working conditions, such as flux densities, frequencies, temperatures, exciting waveforms and dc biases. Core losses are comprised with hysteresis losses and dynamic losses, and a core loss predicted method based on data-driven physical model is proposed. To a certain predicted point, two paths in procedure can be run automatically. First, if there exists two points having the similar working conditions with the predicted point in database, core losses can be predicted directly based on the separated equations. Second, if there are absent loss data for specified waveforms in database, hysteresis losses and dynamic losses under sinusoidal excitation can be used to predict core losses indirectly. Based on the proposed method, the effect of temperature on hysteresis losses and dynamic losses also can be fitted and analyzed individually. The proposed concepts show a good flexibility to handle various exciting waveforms in power electronics and have been verified correctly based on the loss data of 3F3 and 3C90 in MagNet database.

I. INTRODUCTION

Ferrite materials are widely used in designing high frequency magnetic components, which are one of the most important components in power converters. To design a reasonable heat sink, we need to estimate the core loss under different working conditions. Magnetic devices are often subjected to non-sinusoidal waveforms with variable duty cycles, flux densities, temperatures and dc biases. 2023 MagNet Challenge aims to upgrade the Steinmetz equation with the support of a massive amount of measurement data covering different materials across a wide range of frequencies, waveform shapes, temperatures, and dc-bias.

Two major methods stand out in quantifying power losses in soft magnetic materials, which we categorize here as the math-based approach and the physics-based approach. In the former one, the most popular method is using the modified Steinmetz equations. However, the Steinmetz coefficients varies with different temperatures and materials significantly, and they are not usually given by manufacturers and hard to measure. The model based on physical principles has the advantage of predicting the core loss in different conditions. Yet many parameters used in previous physics-based methods are not easily available [1].

In this proposal, we show how to build a predicted model based on the physical mechanisms and the massive data in MagNet database.

II. PHYSICAL MECHANISM OF FERRITE CORE LOSS

According to the modern physicist research, the ferrite core loss can be separated into three main contributions: hysteresis loss, classical eddy current loss and excess eddy current loss [2, 3]. In order to simplify the analysis, we treat that core losses are composed by hysteresis losses and dynamic losses mainly.

Hysteresis losses: hysteresis loss is introduced by the steady-state loss of the Weiss domains. A general format of the hysteresis loss power density can be written as

$$P_h \propto c_h B_p^x f \quad (1)$$

where c_h is the coefficient of the hysteresis loss. When the core is applied by low excitation, the exponent x of B_p is 3. x will change from 3 to 2 at a higher flux level for large excitation. Clearly, the coefficient x changes with the flux level and gives a trouble to model the hysteresis loss. Here, x sets as 2 since power ferrite cores are worked under large flux density normally.

Dynamic losses: The added AC voltage will produce a changing magnetic field and electric field in ferrite materials. This will generate eddy current losses, dielectric losses and other types of losses inside the core. Here, we define all losses associating with the exciting waveforms are dynamic losses. Since the main part of dynamic losses in the using ranges of ferrite materials is the eddy current loss, a general format of the dynamic loss power density can be expressed as

$$P_d \propto c_d B_p^2 f^2 \quad (2)$$

where c_d is the coefficient of the dynamic loss.

Separation equation: For the same ferrite material, c_h in (1) and c_d in (2) can be treated as constant values roughly if the flux density and frequency in two points f_A and f_B are close with each other. Therefore, based on (1) and (2), the components of hysteresis loss power density and dynamic loss power density can be separated [4].

$$\begin{cases} P_v(f_A) = P_h(f_A) + P_d(f_A) \\ P_v(f_B) = P_h(f_B) + P_d(f_B) \end{cases} \Rightarrow \begin{bmatrix} P_v(f_A) \\ P_v(f_B) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \frac{f_B}{f_A} \left(\frac{B_B}{B_A} \right)^2 & \left(\frac{f_B}{f_A} \frac{B_B}{B_A} \right)^2 \end{bmatrix} \begin{bmatrix} P_h(f_A) \\ P_d(f_A) \end{bmatrix} \quad (3)$$

If the predicted point in f_C also has the similar values of the flux density and frequency with f_A and f_B , using (4) to predict core losses will not bring obvious errors.

$$P_v(f_C) = \frac{f_C}{f_A} \left(\frac{B_C}{B_A} \right)^2 P_h(f_A) + \left(\frac{f_C}{f_A} \right)^2 \left(\frac{B_C}{B_A} \right)^2 P_d(f_A) \quad (4)$$

Waveforms: Because ferrite cores have an equivalent R_{ac} , eddy current loss is introduced by the changing magnetic fields inside the core, and their values highly depend on exciting voltage waveforms [2, 3].

$$P_e \propto \frac{1}{R_{ac}} \frac{1}{T} \int_0^T v^2(t) dt \propto \frac{1}{R_{ac}} \frac{1}{T} \int_0^T \left[\frac{dB}{dt} \right]^2 dt \quad (5)$$

where $v(t)$ is the excitation over one switching period. Since dynamic loss power density P_d are also associated with exciting waveforms, the equivalent R_{ac} of dynamic losses under different exciting voltage waveforms can be determined based on the experimental results. In [6], without calculating R_{ac} , core losses under nonsinusoidal waveforms can be predicted directly based on the mathematical relationships with sinusoidal losses.

To the excitation with extreme duty cycles or with a zero voltage period, dielectric losses increase obviously [7]. If dynamic losses are derived based on sinusoidal losses, a modified coefficient $c(D)$ can be multiplied with P_d to embody this additional effect of dielectric losses.

The effect of temperature and dc bias: The coefficients of c_h and c_d in ferrite material show the significant variations with the temperature and dc bias. Benefited from massive data in MagNet database, the change tendencies of hysteresis loss and the dynamic loss under different temperatures and dc biases can be fitted and analyzed easily.

III. FLOWCHART FOR PREDICTING CORE LOSSES

From (1) to (5), the predict point under different combinations of frequencies, temperatures, dc biases and waveforms can be predicted.

Case 1: Absent sinusoidal losses under the certain combinations of f and B in database.

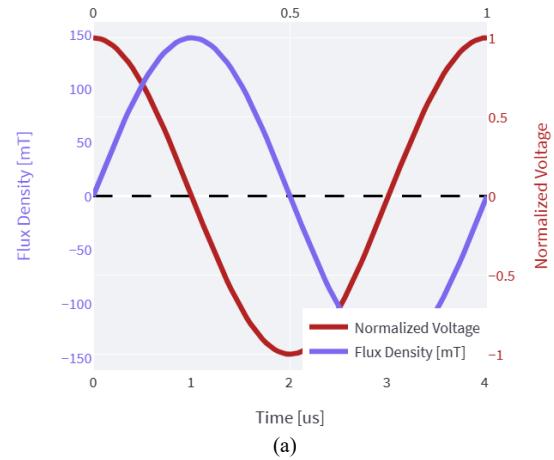
In most cases, predicted points may have different frequencies and flux densities with those existing points in MagNet database. So the procedure should have an ability to predict core losses from those existing points, and the demo waveforms and flowchart are shown in Fig. 1.

Step 1: Selecting suitable data from database. As mentioned before, in order to alleviate the variations of c_h and c_d , points of flux density and frequency in f_A and f_B should select the closest values with the predicted point in f_C .

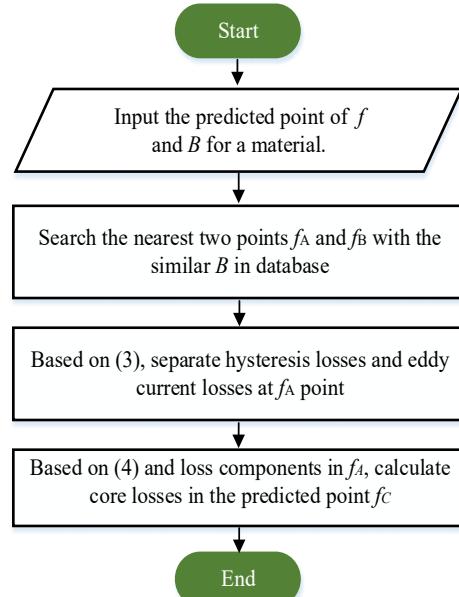
Step 2: Separating hysteresis losses and dynamic losses. (3) can be used to separate hysteresis losses and dynamic losses from two points in f_A and f_B , so the loss component in f_A under certain conditions can be calculated easily.

Step 3: Calculating core losses at the predicted point. Based on (5), core losses in the predicted point f_C can be derived.

Actually, this predicted method also can be available for other types of exciting waveforms directly, such as rectangular excitations or pulse excitations when duty cycles of waveforms equal.



(a)



(b)
Fig. 1 Calculation sinusoidal losses under a random combination of f and B

Case 2: Absent sinusoidal losses under the certain temperature in database.

In this case, the predicted point has a temperature that does not exist in database, and the flowchart is shown in Fig. 2.

Step 1: Calculating hysteresis losses and dynamic losses under the frequency and flux density of the predicted point for typical temperatures in database, as listed in (1).

Step 2: Fitting the changing curves of hysteresis losses and eddy current losses under the typical temperature individually.

Step 3: Calculating the losses components of P_h and P_d under the certain temperature, and the total core loss power density can be derived.

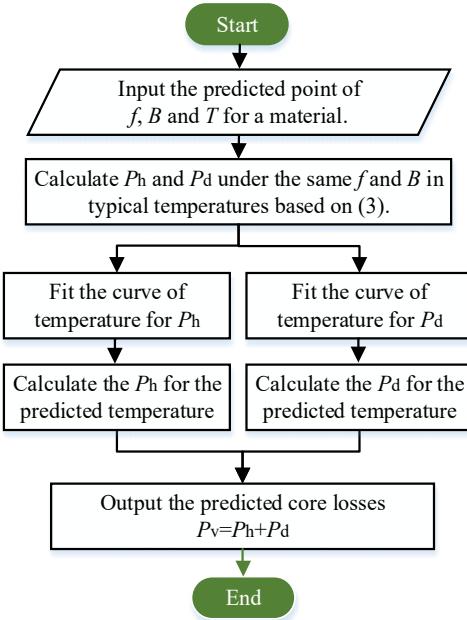


Fig. 2 Calculation sinusoidal losses under a specified temperature

Case 3: Absent rectangular losses under the certain combinations of f and B in database.

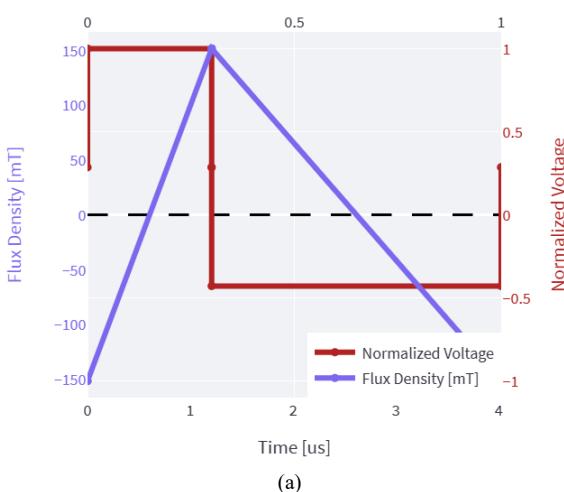
If the flux density and frequency are the same, different voltage waveforms only effect the component of dynamic losses, and will not effect hysteresis losses. In this case, it shows that core losses under rectangular excitations can be derived from sinusoidal losses directly. The demo waveforms and flowchart are shown in Fig. 3.

Step 1: According to the f and B at the predicted point under rectangular excitation, P_h and P_d with the same f and B under sinusoidal excitations are separated.

Step 2: Building the mathematical relationships of dynamic losses between the sinusoidal waveform and the rectangular waveform. P_d under different duty cycles can be calculated, as listed in (6).

$$P_d(D) = \frac{1}{4} \left(\frac{1}{D} + \frac{1}{1-D} \right) \frac{8}{\pi^2} P_d(\sin) \quad (6)$$

Step 3: Calculating the total core losses power density P_v based on P_h and P_d .



(a)

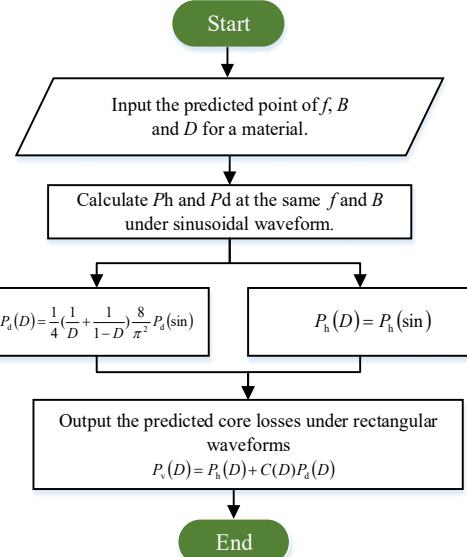


Fig. 3 Calculation rectangular losses from sinusoidal losses

Case 4: Absent pulse losses under the certain combinations of f and B in database.

Similar to the case 3, core losses under pulse excitations also can be derived from sinusoidal losses. In MagNet challenge database, the arbitrary pulse waveforms have four periods which are D_1 , D_2 , D_3 and D_4 , as shown in Fig. 4 (a). The typical amplitudes of voltage in each period are listed in Table I. The flowchart is shown in Fig. 4 (b).

Step 1: According to the f and B at the predicted point under pulse excitations, P_h and P_d at the same f and B under sinusoidal excitations are separated.

Step 2: Building the mathematical relationships of dynamic losses between the sinusoidal waveform and the pulse waveform. P_d under different duty cycles can be calculated. Although pulse waveforms in MagNet challenge show various voltage levels under different duty cycles, the volt-second balance is met. We build the mathematical relationships of dynamic losses between the sinusoidal waveform and the square waveform with $D=0.5$ at first, then $P_d(D)$ under symmetrical pulse waveform with $D=0.4$ can be predicted [7].

$$P_d(D_1 = 0.4) = 1.25 \frac{8}{\pi^2} P_d(\sin) \quad (7)$$

Step 3: Calculating the total core loss power density P_v based on P_h and P_d .

Table I The features of pulse waveforms with positive D_1

D_1	D_2	D_3	D_4	V_{D1}^*	V_{D2}^*	V_{D3}^*	V_{D4}^*
0.1	0.4	0.1	0.4	1	0	-1	0
0.2	0.3	0.2	0.3	1	0	-1	0
0.3	0.3	0.1	0.3	0.667	-0.167	-1	-0.167
0.3	0.2	0.3	0.2	1	0	-1	0
0.4	0.2	0.2	0.2	0.667	-0.167	-1	-0.167
0.4	0.1	0.4	0.1	1	0	-1	0
0.5	0.2	0.1	0.2	0.429	-0.286	-1	-0.286
0.5	0.1	0.3	0.1	0.667	-0.167	-1	-0.167
0.6	0.1	0.2	0.1	0.429	-0.286	-1	-0.286
0.7	0.1	0.1	0.1	0.25	-0.375	-1	-0.375

Taken the amplitude of symmetrical voltage at $D_1 = 0.4$ as a reference. According to the principle of vol-second balance, the voltage relationships meet (8).

$$\begin{cases} V_{D1} = V_{D=0.4} \times \frac{0.4}{D_1}, & V_{D2} = V_{D1} \times \frac{V_{D2}^*}{V_{D1}^*} \\ V_{D3} = V_{D1} \times \frac{V_{D3}^*}{V_{D1}^*}, & V_{D4} = V_{D1} \times \frac{V_{D4}^*}{V_{D1}^*} \end{cases} \quad (8)$$

When duty cycles change, dynamic loss power density can be determined based on the features of pulse waveforms, and their relationships meet (9).

$$\begin{cases} P_d(D_1 = 0.4) = \frac{0.8V_{D=0.4}^2}{R_e} \\ P_d(D) = \frac{V_{D1}^2 D_1 + V_{D2}^2 D_2 + V_{D3}^2 D_3 + V_{D4}^2 D_4}{R_e} \end{cases} \quad (9)$$

Based on (7), (8) and (9), dynamic loss power density under various duty cycles can be calculated.

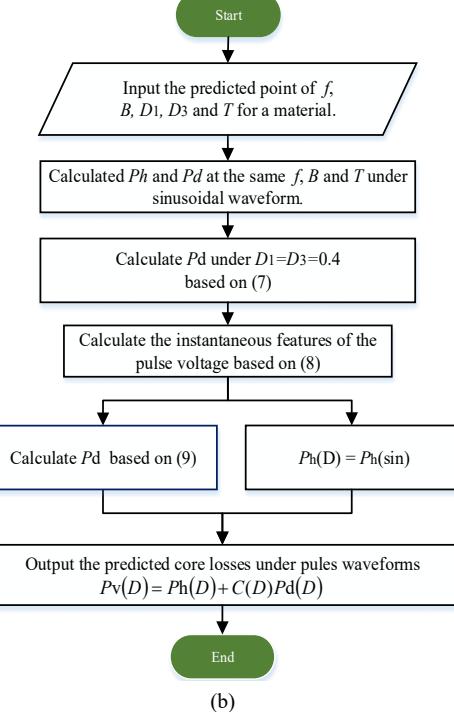
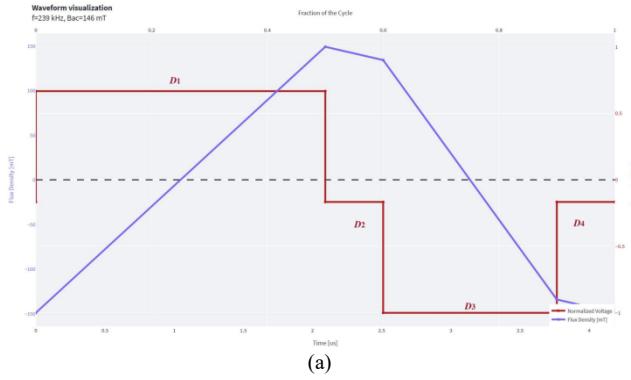


Fig. 4 Calculation pulse losses from sinusoidal losses

IV. RESULTS ANALYSIS

For sinusoidal losses in case 1, Fig. 5 shows the errors between the predicted values and the actual values for N87 @25°C under sinusoidal excitations. The hollow circle represents the actual losses, the star symbol represents the predicted losses, and the solid square represents the errors between the actual losses and the predicted values. Although

the actual points have different combinations of flux density and frequency, the predicted values based on (3) and (4) show a good accuracy and the maximum error in the selected points is 1.5% only.

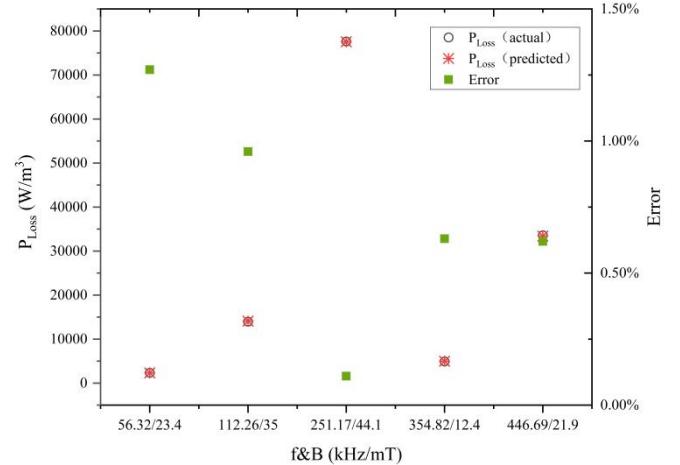


Fig. 5 The predicted values for N87@25°

For the effect of temperature in case 2, N49 is selected as an example. Fig. 6 show the changing tendency of core loss power density, hysteresis loss power density and dynamic loss power density under typical temperatures respectively. Under different combinations of flux density, core loss power density changes nonlinear greatly with temperature, as shown in Fig. 6 (a). This will bring a significant difficulty to predict core losses. Based on (3), hysteresis losses and dynamic losses under typical temperature can be separated individually and used to fit curves, as shown in Fig. 6 (b) and Fig. 6 (c). Therefore, to those points do not exist in database, core loss power density under various temperatures can be predicted simply.

For predicting rectangular losses from sinusoidal losses in case 3, three combinations of flux density and frequency in N87 are selected to predict core loss power density under duty cycles from 0.1 to 0.9, as shown in Fig. 7. Lines are the fitting curves based on rectangular losses in MagNet database, and solid points are predicted values from sinusoidal losses. As mentioned above, dielectric losses can not be neglected under rectangular excitations with extreme duty cycle. Therefore, we select the modified coefficient $c(D) = 1.3$ to multiple the original P_d under $D=0.1$ or $D=0.9$.

For predicting pulse losses from sinusoidal losses in case 3, three combinations of flux density and frequency for N49 are selected to predict core losses under different duty cycles, as shown in Fig. 8. Lines are the fitting curves based on pulse losses in MagNet database, and solid points are predicted values from sinusoidal losses. Again, dielectric losses can not be neglected when pulse waveforms contain zero voltage periods. Therefore, we select the modified coefficient $c(D)=1.48@D=0.1$, $c(D)=1.38@D=0.2$, $c(D)=1.26@D=0.3$ to multiple the original P_d for symmetrical pulse waveforms respectively.

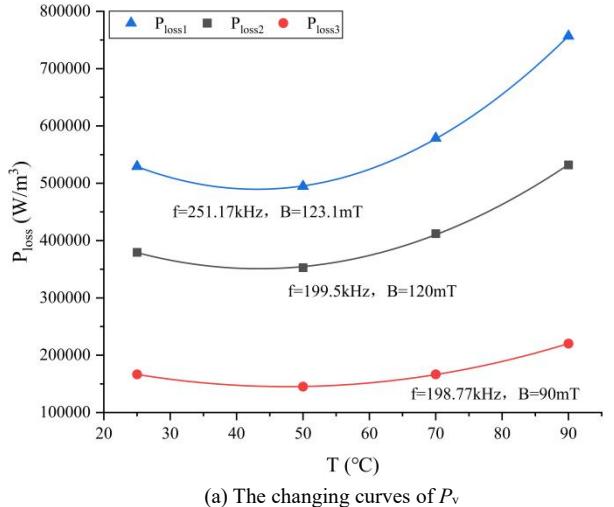
In Table II, the total number of model parameters and model size are listed for the final five unknown materials.

Table II The features of pulse waveforms

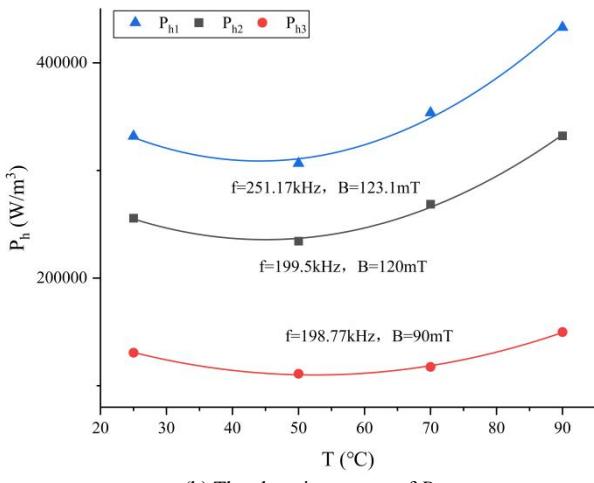
Sample	Numbers of parameter	Size (KB)
1	3 (f, B, T)	51
2	3 (f, B, T)	40
3	3 (f, B, T)	49
4	3 (f, B, T)	60
5	3 (f, B, T)	49

V. CONCLUSION

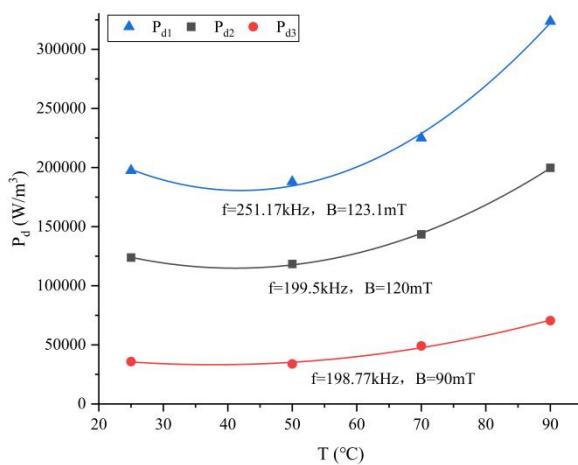
In this paper, core losses are separated into hysteresis losses and dynamic losses, and a data-driven physical prediction model of core losses is proposed. Based on losses in MagNet database, core losses under arbitrary excitations can be predicted directly. At the same time, the effect of temperature on hysteresis losses and dynamic losses also can be fitted and analyzed individually. The proposed concepts show a good flexibility to predict core losses under various exciting waveforms in power electronics.



(a) The changing curves of P_v



(b) The changing curves of P_h



(c) The changing curves of P_d

Fig. 6 The changing tendency under different temperatures for N49

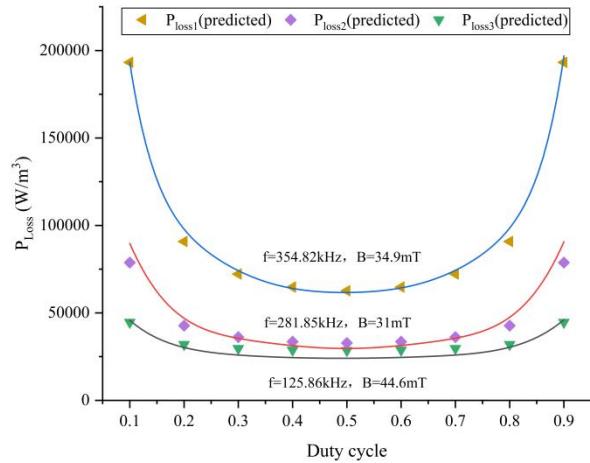


Fig. 7 The predicted values under rectangular excitation for N87@25°

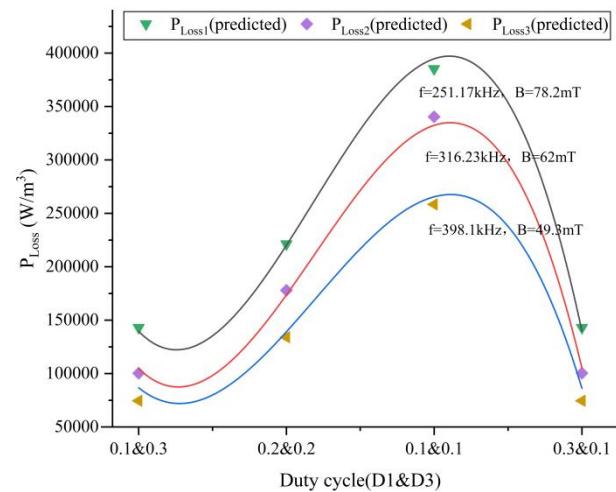


Fig. 8 The changing tendency under pulse waveform for N49

REFERENCES

- [1] Bertotti G, "General properties of power loss in soft magnetic material," *IEEE Trans. Magn.*, vol.24, no.1, pp.621-630, Jan.1988.
- [2] W. A.Roshen, "A practical, accurate and very general core loss model for nonsinusoidal waveforms," *IEEE Trans. Power Electro.*, vol.22, no.1, pp. 30–40, Jan.2007.
- [3] W. A.Roshen, "Ferrite core loss for power magnetic components design," *IEEE Trans. Magn.*, vol.27, no.6, pp.4407-4415, Nov.1991.
- [4] Z. Yan, Sun ai-ming, "A Simplified Ferrite Core Loss Separation Model for Switched Mode Power Converter," *IET Power Electro.*, vol.9, no.3, pp.529-535, Mar.2016.
- [5] Z. Yan, Chen Qimi, and Zhang Junbo, "Predicting Core Losses Under the dc Bias Based on the Separation Model," *IEEE Journal of Emerging and Selected Topics in Power Electro.*, vol. 5, no. 2, pp. 833–840, June. 2017.
- [6] Z. Yan, Z. Weibo and T. Guanghui, "A Core Loss Calculation Method for DC/DC Power Converters Based on Sinusoidal Losses," in *IEEE Transactions on Power Electronics*, vol. 38, no. 1, pp. 692-702, Jan. 2023.
- [7] Z. Yan, W. Xie, C. Ma and L. Huang, "Modeling Core Losses Under Exciting Waveforms With Zero Voltage Period in Power Converters," in *IEEE Transactions on Industrial Electronics*, vol. 70, no. 2, pp. 1378-1386, Feb. 2023.

Attention Is Magnetic Core Loss Modelling Need?

Yaohua Li, *Student Member, IEEE*, Sicheng Wang, *Student Member, IEEE*, Yue Wu, *Student Member, IEEE*, Yongbin Jiang, *Member, IEEE*, Ziheng Xiao, *Member, IEEE*, and Yi Tang, *Senior Member, IEEE*

Abstract—Given the complex mechanism of magnetic core behaviors, the traditional equation-based models are usually difficult to achieve high accuracy under various excitations and operating conditions. In this report, we explore the potential of state-of-the-art transformer models in magnetic core loss modelling. The result shows that the proposed CNN-transformer model can achieve a higher accuracy than the existed transformer model without sacrificing the computational efficiency and parameter size. However, the attention-based transformer model needs abundant training data to achieve a good generalization ability and predicted performance. Therefore, for a small dataset, the attention mechanism may not be a good choice for magnetic core loss modelling.

Index Terms—Magnetic core loss modelling, Transformer, CNN, MagNet.

I. INTRODUCTION

DEEP learning provides a powerful tool for magnetic core loss modelling. Compared with traditional equation-based models, data-driven models can achieve higher accuracy and better generalization ability in consideration of sufficient training data and model parameters.

Deep learning-based prediction and recognition technologies commonly utilize Convolutional Neural Networks (CNNs) for the purpose of feature learning and classification. The use of convolution operations in CNN models facilitates the sharing of parameters, thereby reducing the total number of parameters that require training. This effectively diminishes the complexity of the model. CNNs possess the capability to autonomously learn representations of input data features, which significantly lightens the load of manual feature extraction. Nonetheless, when dealing with prediction or recognition subjects with intricate features, it becomes essential to augment the number of convolution layers in the CNN to grasp more sophisticated and advanced features. This, however, results in increased computational expenses.

Over the past few years, Transformer models have made significant strides in both natural language processing and computer vision. These models excel at identifying long-term dependencies in input sequences, facilitating comprehensive global feature extraction. The inherent parallel processing capability of Transformers enhances their efficiency in handling input sequences, thereby accelerating both training and inference processes. However, their reliance on self-attention mechanisms results in a reduced capacity for in-depth feature exploration, which in turn elevates the model's complexity and necessitates higher quality input data.

Due to the highly nonlinear characteristics of magnetic materials, we adopt a strategy that combines Transformers and CNNs to compensate for the drawbacks of both mainstream

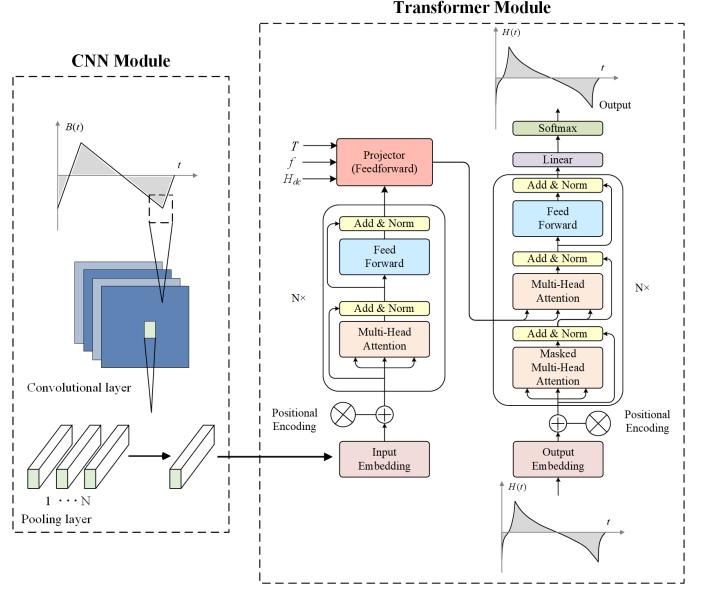


Fig. 1. Proposed CNN-Transformer structure.

deep learning models. Specifically, a CNN layer is used as the regression head, while the Transformer serves as the backbone of the model. This modeling approach takes into account the complementary nature of global and local information, which helps to significantly reduce training time while enhancing the model's multi-scale feature extraction capabilities.

II. MODEL ARCHITECTURES

During the entire Magnet Challenge process, we experimented with the following types of models.

A. Vision Transformer Model

The Vision Transformer (ViT) is essentially tailored for visual tasks, utilizing a Transformer-based architecture. It operates on images by breaking them down into numerous small pieces, referred to as "patches," and then handling these patches in a sequential manner. This approach is same as how traditional Transformers process textual data, considering each word in a sequence for natural language tasks. In this case, the input magnetic flux density $B(t)$ is treated as a one-dimensional(1-D) image for ViT, where it undergoes patch embedding before being fed through a Transformer structure, enabling the prediction of the magnetic field intensity sequence $H(t)$. However, our experiments indicate that ViT's head is relatively intricate, and it only performs comparably to a stand-alone Transformer when provided with sufficient input data.

TABLE I
COMPARISON OF DIFFERENT MODELS WITH BASELINE

Name	Model	Parameters	CNN In	Transformer In	Decoder Layer	Average Error	95-Prct Error	Running Time
Baseline	Transformer	28481	NA	128*1	1	3.05%	9.64%	129mins
Model 1	Transformer	28481	NA	256*1	1	2.47%	8.38%	304mins
Model 2	CNN+Transformer	28564	256	128*2	1	2.4%	7.94%	152mins
Model 3	CNN+Transformer	28564	512	128*2	1	2.07%	6.97%	152mins
Model 4	CNN+Transformer	62788	512	128*2	4	1.45%	4.72%	502mins

Furthermore, segmenting the input into multiple patches can disrupt the inherent relationships within the input sequences. Adequate parameters are crucial to ensure sufficient feature learning, which is necessary to reduce prediction errors.

B. Long Short-Term Memory Network Model

LSTM stands out as an apt solution when dealing with shorter data sequences or in environments with constrained resources. Yet, in light of the complex nonlinear attributes of magnetic materials, LSTM's capacity for parallel processing is markedly less effective compared to that of transformers. Moreover, transformers present a broader horizon for improvements. As the database expands with increasing amounts of data, the advantages of transformers will become more distinctly observable.

C. CNN-Transformer Model

Based on our experience with ViT experiments, our next consideration is how to replace the head of ViT with a more lightweight structure that can better extract features from the input sequence. Undoubtedly, CNN offers such a capability. The input sequence can first be transformed into a multi-dimensional sequence through CNN before entering the transformer. This approach not only allows for the input of more sampling points to enhance the model's prediction accuracy but also reduces the model's runtime.

III. EXPERIMENTAL RESULTS

Tables I and Fig. 2 illustrate the training performance of five different models on the magnetic material N87. All models are tested on the same GPU(Nvidia RTX 4090). The Baseline model, the Transformer provided from the MagNet Challenge tutorial, stands out for its shortest training time, showcasing the parallel processing advantage of Transformers. However, this model also exhibits the highest error, likely due to Transformers' dependency on extensive and high-quality labeled data, a limitation in certain applications and tasks. Model 1 indicates that expanding the number of sequences fed into the transformer is able to increase the accuracy of predicted core loss, though at the cost of doubling the runtime. This led us to explore balancing enhanced accuracy with reduced runtime. Model 2 reveals that incorporating an additional CNN layer not only furthers accuracy improvement but also substantially cuts down simulation time, nearing Baseline levels. Leveraging CNN's robust feature extraction, Model 3 increases input size from 256 to 512, resulting in a further drop in the 95th

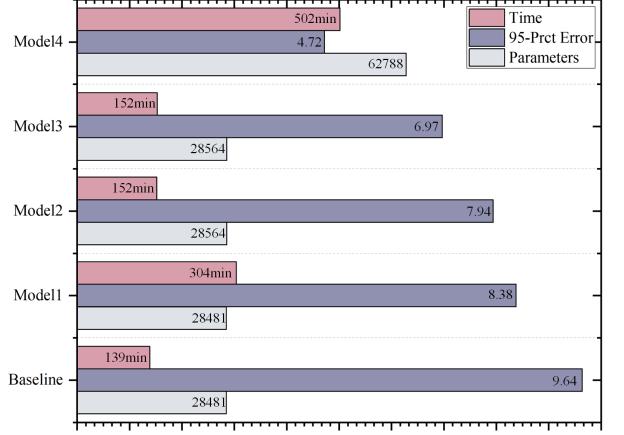


Fig. 2. Comparison of runtime, 95-Prct error and parameters for five models.

percentile error while maintaining the same runtime as Model 2. In summary, Model 3 optimally demonstrates the advantages of a combined model approach, aligning well with our goals.

To push for even lower error rates, Model 4, an extension of Model 3, upscales the parameter count from 28,564 to 62,788. Comparative data from Fig. 2 clearly shows that a higher parameter count markedly lowers the error rate, but also leads to a significant increase in computational time.

IV. DISCUSSION

A. Model Size and Model Parameters

The model size and model parameters are two important factors that affect the performance of the model. For attention-based transformer model, the model size is crucial than CNN or LSTM model. Based on our experimental results, the model size of transformer should be no less than 10000 and the parameters should be around 30000. In this case, the accuracy and runtime of the model can be guaranteed.

B. Transfer Learning Ability

Although Transformer models excel in many aspects, their success in transfer learning largely depends on factors such as the amount of data, the similarity between materials, and model adjustment strategies. The low performance of our

model in transfer learning can be attributed to several reasons. Firstly, CNN-Transformer model has a strong dependency on data. In transfer learning scenarios with limited data, Transformers may struggle to effectively learn key features of the new magnetic material, leading to decreased performance. Secondly, when there is a significant difference in data distribution or features between the source and target tasks, Transformers might find it challenging to transfer the learned knowledge effectively. This issue arises because their self-attention mechanism might overly focus on specific patterns from the source task that are irrelevant or misleading for the new task.

Moreover, due to their large parameter count, Transformer models are prone to overfitting on limited datasets, making it difficult to adapt to new tasks. Furthermore, effective transfer learning involves not just the transfer of model parameters but also requires fine-tuning according to the characteristics of the target task. Without targeted adjustments, Transformers might not perform optimally on the new task.

V. CONCLUSION

In conclusion, this report has explored the effectiveness of advanced transformer models in the realm of magnetic core loss modeling, addressing the limitations of traditional equation-based models that struggle with accuracy under diverse excitations and operating conditions. The results clearly demonstrate that the proposed CNN-transformer model outperforms existing transformer models in terms of accuracy, while maintaining computational efficiency and a manageable parameter size. However, it is important to note that the attention-based transformer model, despite its strengths, requires a substantial amount of training data to develop robust generalization capabilities and achieve high predictive performance. Consequently, when dealing with smaller datasets, the attention mechanism might not be the most suitable approach for magnetic core loss modeling. This insight is crucial for guiding future research and application of machine learning techniques in magnetic core modeling, especially in scenarios where data availability is limited.

ACKNOWLEDGMENTS

We would like to express our gratitude to organizers for organizing the MagNet Challenge 2023 and providing valuable equation-based insights and machine learning based solutions. We sincerely hope that through our globally collective efforts, we can make advancements in the modelling of magnetic losses. Thank you for this opportunity to collaborate and work towards furthering our understanding in this field.

MagNet Challenge 2023 Final Report

*Yun-Shan Hsieh, *Jian-De Li, *Li-Chen Yu, *Tzu-Chieh Hsu, *Yu-Chen Liu, Senior Member, IEEE,
 +Chin-Hsien Hsia, Senior Member, IEEE, § Chen Chen, Member, IEEE

*National Taipei University of Technology, Taiwan

+National Ilan University, Taiwan

§LITEON Technology, Taiwan

Abstract- In this report, a novel architecture of a Feedforward Neural Network (FNN) is proposed for the prediction of core loss. The architecture was refined using Optuna, resulting in a network configuration with five hidden layers, consisting of 65, 55, 116, 40, and 123 neurons, respectively. The model's hyperparameters were optimized, featuring a decay epoch of 423, a decay ratio of 0.458, and an initial learning rate set at 0.005. To enhance the model's precision and operational efficiency, Z-score normalization techniques were employed. Addressing the challenge of scarce training data, the study integrated transfer learning with the Dynamic Time Warping (DTW) algorithm. This integration facilitated the evaluation of the correlation between 10 pre-test materials and an unknown material, leading to the selection of the most pertinent materials for pre-training. It was observed that pre-training did not significantly improve the results for Materials A and B, thereby restricting the application of transfer learning to Materials C, D, and E. The dataset was divided into 60% for training, 20% for validation, and 20% for testing purposes. The results indicated that all materials, with the exception of Material D, achieved an average error rate below 5%.

I. INTRODUCTION

Magnetic components in converters, which often operate under high-frequency conditions, are subject to substantial variations in their excitation sources. These sources can encompass a range of waveforms, including square waves, triangular waves, and those incorporating a DC component. Traditionally, the calculation of magnetic core losses in such environments has predominantly relied on the Steinmetz Equation (SE) or the Original Steinmetz Equation (OSE) [1]. These formulas are primarily designed for sine wave excitations that do not involve a DC bias, as depicted in Equation (1). This conventional approach, while effective under specific conditions, may not fully encapsulate the complexities encountered in varied waveform scenarios.

$$P_{SE} = kf^\alpha B_S^\beta \quad (1)$$

However, in switch-mode power supplies, sine waves are typically not used, and the excitation source often appears in the form of triangular waves or square waves. Different excitation sources generate different magnetic hysteresis loops. If the same loss calculation method is still used, significant errors will be introduced. Therefore, from the SE, various calculation equations have been derived to accommodate non-sinusoidal excitations and suit different scenarios. These include the Generalized Steinmetz Equation (GSE) [2], Improved Generalized Steinmetz Equation (iGSE) [3], I2 Generalized Steinmetz Equation (i²GSE) [4], and Equivalent Elliptical Loop (EEL) [5]. Unfortunately, these loss calculation methods are not

able to fully capture the differences in different temperatures or waveforms with a bias current. Therefore, there is still room for improvement in the current loss equations.

The emergence of neural networks has ushered in a transformative phase in machine learning, with deep neural networks (DNNs) standing out as a particularly revolutionary advancement. Exhibiting exceptional aptitude for tackling complex, nonlinear issues, which are commonplace in multivariate classification and regression, these networks are distinguished by their sophisticated architectures and an innate ability to identify intricate patterns. This aptitude renders DNNs exceptionally suitable for challenging tasks such as core loss prediction, characterized by its nonlinear and multi-dimensional nature. Our report is dedicated to elaborating on the utilization of deep neural networks for the precise prediction of core losses.

Section 2 delves into the detailed architecture of our neural model and the strategies employed for data processing. Employing a conventional Feedforward Neural Network (FNN) structure, our model has been subjected to meticulous fine-tuning, ensuring its optimal performance. This process of fine-tuning encompasses a strategic calibration of the network's layers and hyperparameters, directed by the Optuna optimization framework. In Section 3, we will discuss the application of transfer learning techniques, emphasizing their utility in surmounting the unique data distribution challenges presented in this study. Section 4 is committed to a comprehensive analysis of our findings, including a thorough comparison with established benchmarks in the Pre-test Report. The report will culminate in Section 5, providing a complete synthesis of our research outcomes. Lastly, Section 6 will outline potential directions for future research and enhancements in this field.

II. NEURAL MODEL ARCHITECTURE AND DATA PROCESSING

This section will explain the design of the models' structures and the concepts of data is processed.

A. Feed-forward Neural Network Core Loss Model

Feed-forward Neural Networks (FNNs) are among the most basic and commonly employed types of artificial neural networks, demonstrating their effectiveness in addressing complex problems involving multiple variables and nonlinear regression. Figure 1 shows the structure of our model. The input layer takes in three types of parameters: frequency 'f', magnetic flux density 'B₀' to 'B₁₀₂₃' for each point in a cycle, and

temperature ' T '. The output layer produces one parameter: the loss per unit volume ' P_{cv} '. In between, the data passes through ' n ' hidden layers, each with a different number of neurons. The exact numbers of neurons in these layers are determined using the Optuna[999] tool, which will be explained in more detail in the following sections.

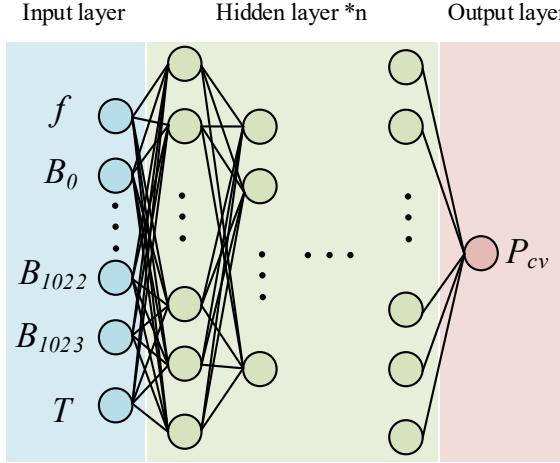


Fig. 1 A schematic diagram of the FNN architecture.

This model is synthesized and trained using PyTorch. It employs ReLU as the activation function to learn non-linearity and uses mean squared error (MSE) of logarithmic core loss as the loss function to ensure uniform performance across the entire operational range. The training optimizer selected is Adam, with an exponentially decaying learning rate strategy implemented for better model convergence. Additionally, the data is split into 60% for training, 20% for validation, and 20% for testing to assess the model's performance.

B. Optuna

Hyperparameter tuning is one of the most tedious tasks in machine learning projects. With the widespread adoption of deep learning methods, its complexity continues to increase, and the need for an effective automated hyperparameter tuning framework is more urgent than ever. To meet the demand for finding better hyperparameter configurations, our team has utilized the Optuna tool.

Optuna is an open-source Python library designed for hyperparameter optimization. Its principles are based on techniques such as Bayesian Optimization and Tree-structured Parzen Estimator (TPE) [8]. Bayesian Optimization involves constructing a probability model for the objective function and then selecting points in the search space that are most likely to improve performance. TPE is a variant of Bayesian Optimization that efficiently explores and exploits the search space using a tree-like structure.

Optuna uses Bayesian Optimization to select the optimal hyperparameter combination by defining the objective function, search space, and evaluation criteria. In addition to this, Optuna provides various built-in optimization algorithms and visualization tools, allowing us to easily conduct experiments and analysis for hyperparameter optimization. This capability

has enabled us to achieve optimal performance in the project focused on predicting iron core losses.

Table.1 presents the custom-defined model hidden layers and neurons, and the configurations identified using Optuna. Table.2 compares the average percentage errors in predicting the iron core losses for ten different materials between these two configurations.

Table.1 A comparison chart of different model settings.

Model setting	Customize	Optuna
Number of hidden layers	3	5
Number of neurons in each hidden layer	15,15,15	65,55,116,40,123

Table.2 Prediction of average percentage error in core losses for ten types of materials using different models (%).

Material	3C90	3C94	3E6	3F4	77
Customize	7.28	6.1	9.48	10.31	7.16
Optuna	4.33	5.19	7.21	8.52	4.17
Material	78	N27	N30	N49	N87
Customize	7.36	9.23	5.62	7.27	11.51
Optuna	3.31	4	3.52	6.63	7.46

After obtaining the results of using Optuna to search for the number of hidden layers and neuron counts, we observed significant differences in the average percentage errors for iron core losses among ten different materials between the custom-defined and Optuna-searched hyperparameter configurations. Consequently, we further attempted to explore the initial value of the learning rate, the decay period of the learning rate, and the decay ratio of the learning rate. Table.3 presents our original hyperparameter settings alongside those identified using Optuna, with the model layers and neuron counts adopting the configuration found in Table 1 through Optuna. Table.4 compares the configurations of these two settings in predicting the average percentage errors for iron core losses of the final five materials.

Table.3 Hyperparameter settings

Hyperparameter	Customize	Optuna
Epoch	2000	2000
Decay_epoch	300	423
Decay_ratio	0.5	0.458
Initial learning rate	0.02	0.005

Table.4 Prediction of average percentage error in core losses for five types of materials in final using different hyperparameters (%)

Average percentage error	Material A	Material B	Material C	Material D	Material E	AVG
Customize	14.35	3.29	5.94	12.11	10.05	9.15
Optuna	5.57	2.17	2.78	11.61	3.59	5.14

C. Optimizing Data with Z-Score Normalization

Input data is processed using the Z-Score method (as defined in Equation 2, where 'x' represents an individual input value, ' μ ' is the mean of the entire data set, ' σ ' is the standard deviation of the entire data set, and 'Z' is the standardized value). The purpose of this method is to scale input variables to a standard range. This process helps in reducing the scale differences between different features, making algorithms more effective and quicker to converge. It improves the model's accuracy and performance, particularly in datasets with variables of varying scales and units, ensuring that each feature contributes equally to the analysis.

$$Z = \frac{x - \mu}{\sigma} \quad (2)$$

III. TRANSFER LEARNING

This section will explain why use Transfer Learning and provide an introduction to its concepts.

A. Data analysis

Table 5 is the dataset for the final evaluation. There's a noticeable difference in the amount of training and testing data for Materials A and D. Material A has 2432 data points for training compared to 7651 for testing. Material D has a mere 580 points for training while there are 7299 for testing. This large discrepancy could lead to the model not having enough data to learn important features during training, which may negatively impact its performance on data it hasn't seen before.

TABLE 5 The dataset for the final evaluation

Number of Data Points	Material A	Material B	Material C	Material D	Material E
Training Data	2432	7400	5357	580	2013
Testing Data	7651	3172	5357	7299	3738

However, transfer learning allows the model to apply knowledge gained from similar tasks to better adapt to the test set. Pre-trained models are beneficial because they have already learned valuable features from a large dataset, providing a strong starting point even when there is a smaller amount of training data. The key principles of transfer learning methods are illustrated in Fig. 2.

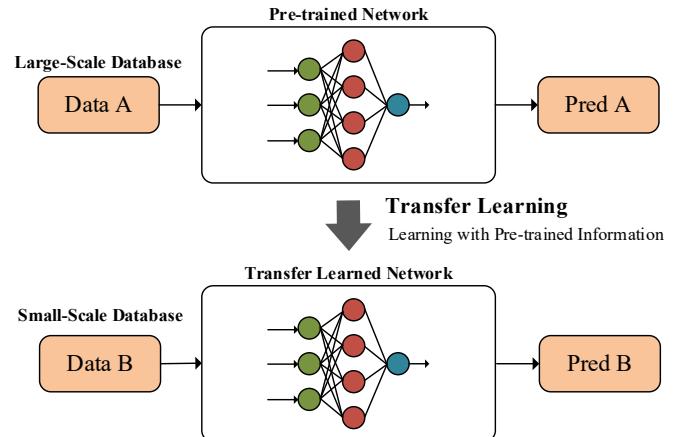


Fig. 2 The process of transfer learning: a network pre-trained on a large database is adapted to make predictions on a smaller database.

B. Decide Pre-train Models

When faced with a limited amount of training data, the filtering of pre-training data becomes crucial. In such scenarios, the model may be at risk of overfitting, relying too heavily on a limited dataset and struggling to generalize to new, similar conditions. To overcome this challenge, careful selection of pre-training data that effectively represents the target application scenarios is essential.

This report presents a method for pre-training data selection by identifying B H curves that closely resemble those of the target training model under specific conditions of temperature, frequency, and loss. The recognition method employs the Dynamic Time Warping (DTW) algorithm [9], which effectively calculates the similarity between two B H curves. As shown in Fig 3, lower DTW values indicate higher waveform similarity.

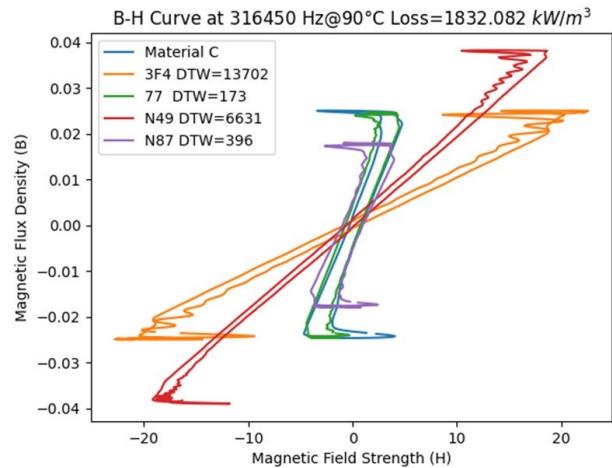


Fig. 3 Under specific conditions, the similarity of the BH curve between Material C and other materials.

According to Table 6, we can clearly observe the mutual relationship between the pre-trained model and the training model. These values are obtained through the DTW algorithm to determine which material's B-H curve is most similar under

specific conditions. Through statistical analysis, we obtain the data in the table. A higher value indicates a higher degree of similarity between the pre-training dataset and the training dataset. This metric serves as a basis for filtering the data used for training.

Table 6 The number of similar BH curve among different materials.

Material	3C90	3C94	3E6	3F4	77
A	75	132	294	1	341
B	93	223	1934	0	792
C	263	386	152	12	728
D	47	11	0	28	3
E	11	1	0	152	5
Material	78	N27	N30	N49	N87
A	256	181	348	5	98
B	528	259	2241	5	217
C	808	620	446	60	444
D	21	37	9	216	26
E	17	30	2	462	6

Ultimately, we opted for the materials with the highest metrics as our pre-training dataset. Thus, Materials A and B will use N30 as the pre-training model, Material C will use 78, and Materials D and E will use N49 as the training model. The architecture of the pre-trained models is as described in Section 2, resulting in three different pre-trained models to be used for training the five materials tested in this competition.

C. Pre-train Result

In this section, we assessed the effectiveness of pre-training across different materials. The findings, detailed in Table 7, indicated that transfer learning significantly benefited Materials C, D, and E, reducing prediction errors. However, for Materials A and B, pre-training did not yield positive results. Material A showed minimal improvement, likely due to its unique characteristics compared to other materials in the dataset. For Material B, which had ample training data, pre-training adversely affected its learning process. Based on these observations, we strategically applied transfer learning only to Materials C, D, and E, while avoiding it for Materials A and B. This decision highlights the need for a tailored approach in applying advanced techniques like transfer learning, depending on each material's specific traits and data availability.

TABLE 7 Comparison of Core Loss Prediction Error with/without Pre-training (%)

Material	Material A	Material B	Material C	Material D	Material E
Without Pretrain	4.635	1.94	3.71	11	4.35
With Pretrain	5.29	2.76	2.65	9.33	3.62

IV. RESULT

In this section, we present a comprehensive analysis of our core loss prediction model's performance. Our evaluation is centered on assessing the accuracy and reliability of the model across different materials, ranging from Material A to Material E. The results are quantified in terms of average error rates, percentiles of relative cumulative frequency, and maximum errors. Additionally, we provide insights into the uniformity of the model's architecture by examining the number of parameters for each material model. This detailed analysis aims to offer a clear understanding of the model's efficacy in predicting core losses under varied conditions.

A. Result

Table 8 presents the error data for Materials A to E, featuring metrics such as the average error, the 95th percentile of relative cumulative frequency (95-PRCT), the 99th percentile (99-PRCT), and the maximum error (Max). Except for Material D, the average errors for the other materials do not exceed 5%. Table 9 lists the Number of Parameters for each material model, which is uniform at 86,728 across all models due to the identical model architecture employed.

TABLE 8 Core Loss Prediction Error

Material	Material A	Material B	Material C	Material D	Material E
Avg	4.635	1.94	2.65	9.33	3.62
95-PRCT	14	4.84	6.95	30.03	12.26
99-PRCT	29.58	12.156	12.69	41.77	22.97
Max	59.56	58.049	49.05	46.6	49.56

TABLE 9 Number of Parameters for Each Material

Material	A	B	C	D	E
Number of Parameters	86728	86728	86728	86728	86728

B. Comparison of Errors Relative to the Pre-test Report

Table 10 presents the progress from midterm to final evaluations across ten materials, showing significant improvements in the final scores. The initial approach during the midterm, which utilized LSTM to forecast the magnetic field intensity H and subsequently calculated losses via the BH curve area, did not yield the expected results. This shortfall might be due to the limited complexity of the LSTM, leading to less precise predictions of H, and the inherent inaccuracies when correlating the BH curve area with actual volumetric core losses. Ultimately, transitioning to a conventional FNN model, although more parameter-intensive, facilitated a more accurate direct prediction of losses than the indirect BH curve area method.

Table.10 Core Loss Prediction Error: Midterm vs. Final Comparison (%).

Material	3C90	3C94	3E6	3F4	77
Pre-test Report	48.58	46.61	23.99	112.10	49.45
Final-test Report	12.2	12.81	16.55	21.2	11.81
Material	78	N27	N30	N49	N87
Pre-test Report	49.45	41.13	19.58	173.50	32.91
Final-test Report	9.31	10.18	8.34	17.23	20.2

The results outlined in this report underscore the effectiveness of our deep neural network model in predicting core losses across a diverse range of materials. Notably, except for Material D, the model maintains an average error rate below 5%, signifying its high accuracy. The uniformity in the number of parameters across all material models, standing at 86,728, reflects a consistent approach in the model's architecture. This consistency further reinforces the reliability of our findings. The data presented in Tables 8 and 9 collectively demonstrate the model's robustness and precision, making it a valuable tool in the field of core loss prediction.

V. CONCLUSION

In conclusion, this report successfully illustrates the effective use of a Feedforward Neural Network (FNN) in predicting core loss. Through the meticulous fine-tuning of the network architecture using Optuna, a robust model comprising five hidden layers with varying neuron counts (65, 55, 116, 40, and 123) was developed. The calibration of hyperparameters, including a decay epoch of 423, a decay ratio of 0.458, and an initial learning rate of 0.005, played a significant role in enhancing the model's performance. The application of Z-score normalization was particularly effective in improving accuracy and efficiency.

Addressing the challenge of limited training data, the incorporation of transfer learning and the Dynamic Time Warping (DTW) algorithm proved instrumental. This approach enabled the identification of the most relevant pre-test materials for pre-training, optimizing the model's learning process. However, it's noteworthy that pre-training did not improve outcomes for Materials A and B, indicating that the effectiveness of transfer learning can vary depending on the material type. Consequently, transfer learning was selectively applied to Materials C, D, and E.

The data partitioning strategy, with 60% allocated for training, 20% for validation, and 20% for testing, was effective in achieving an average error below 5% for most materials. The notable exception was Material D, which demonstrated the model's limitations and potential areas for future improvement. Additionally, Material B exhibited the highest error rate at

58.049%, indicating a need for further refinement in the model's approach to specific material types.

VI . FUTURE WORK

Future research will focus on enhancing accuracy for challenging materials, like the high error rates seen in Material B (58.049%). This involves integrating diverse datasets with a wider range of material characteristics and developing specialized algorithms for materials like Material D.

Given the variable success of transfer learning, future studies will explore sophisticated methods to identify material similarities and extend transfer learning to more materials. Improving pre-training techniques, particularly for less responsive materials, is also a key goal, possibly through advanced feature extraction methods.

To address limited training data, the plan includes expanding the dataset through synthetic data generation or additional real-world data collection. Parallel to this, there will be a continuous effort in optimizing hyperparameters like learning rates and neuron configurations, enhancing model performance and accuracy.

Future models will benefit from integrating multi-modal data sources, such as thermal and electrical properties. Developing ensemble models that combine the FNN with other neural network types may yield better results for challenging materials.

These strategies aim to overcome current limitations, significantly improving the model's accuracy and applicability in core loss prediction, and enhancing its utility in industrial applications.

REFERENCES

- [1] 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, July-Aug. 2001, doi: 10.1109/28.936396
- [2] J. Li, T. Abdallah, and C. R. Sullivan, "Improved Calculation of Core Loss With Nonsinusoidal Waveforms", IEEE Industry Applications Society Annual Meeting, Oct. 2001, pp. 2203-2210.
- [3] K. Venkatachalani, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate Prediction of Ferrite Core Loss with Nonsinusoidal Waveforms Using Only Steinmetz Parameters", 2002 IEEE Workshop on Computers in Power Electronics, 2002, pg. 36-41. Reference 181.
- [4] J. Muhlethaler, J. Biela, J.W. Kolar, A. Ecklebe, "Improved Core-Loss Calculation for Magnetic Components Employed in Power Electronic Systems," Power Electronics, IEEE Transactions on , vol.27, no.2, pp.964-973, Feb. 2012.
- [5] D. Lin, P. Zhou, W. N. Fu, Z. Badics and Z. J. Cendes, "A dynamic core loss model for soft ferromagnetic and power ferrite materials in transient finite element analysis," in IEEE Transactions on Magnetics, vol. 40, no. 2, pp. 1318-1321, March 2004, doi: 10.1109/TMAG.2004.825025.
- [6] J. Mühlethaler, J. Biela, J. W Kolar, and A. Ecklebe, "Core Losses Under the DC Bias Condition Based on Steinmetz Parameters," in IEEE Transactions on Power Electronics, vol. 27, no. 2, pp. 953-963, FEBRUARY 2012, doi: /TPEL_2011_2160971.
- [7] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85-117, jan 2015.
- [8] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a Next-Generation Hyperparameter Optimization Framework," Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.
- [9] M. Müller, "Dynamic time warping," in *Information retrieval for music and motion*, pp. 69-84, 2007.

HARDCORE: H-field and power loss estimation for arbitrary waveforms with residual, dilated convolutional neural networks in ferrite cores

Nikolas Förster*, Wilhelm Kirchgässner*, Till Piepenbrock*, Oliver Schweins*, Oliver Wallscheid*

*Department of Power Electronics and Electrical Drives
Paderborn University, 33095 Paderborn, Germany

Abstract—The MagNet Challenge 2023 calls upon competitors to develop data-driven models for the material-specific, waveform-agnostic estimation of steady-state power losses in toroidal ferrite cores. The following HARDCORE (H-field and power loss estimation for Arbitrary waveforms with Residual, Dilated convolutional neural networks in ferrite COREs) approach shows that a residual convolutional neural network with physics-informed extensions can serve this task efficiently when trained on observational data beforehand. One key solution element is an intermediate model layer which first reconstructs the bh curve and then estimates the power losses based on the curve’s area rendering the proposed topology physically interpretable. In addition, emphasis was placed on expert-based feature engineering and information-rich inputs in order to enable a lean model architecture. A model is trained from scratch for each material, while the topology remains the same. A Pareto-style trade-off between model size and estimation accuracy is demonstrated, which yields an optimum at as low as 1755 parameters and down to below 8 % for the 95-th percentile of the relative error for the worst-case material with sufficient samples.

Index Terms—Magnetics, machine learning, residual model

I. INTRODUCTION

The MagNet Challenge 2023 is tackled with a material-agnostic residual convolutional neural network (CNN) topology with physics-informed extensions in order to leverage domain knowledge. Topological design decisions are dictated by peculiarities found in the data sets and by the overall goal of maximum estimation accuracy at minimum model sizes.

The topology’s central idea is the calculation of the area within the bh polygon based on a preceding h sequence estimate, see Fig. 3. The area within the polygon formed by the sequences $\mathbf{b}, \mathbf{h} \in \mathbb{R}^{1024}$ can be calculated using the shoelace formula or surveyor’s area formula [1]. The shoelace method assigns a trapezoid to each edge of the polygon as depicted in Fig. 1. The area of these trapezoids is defined according to shoelace either with a positive or negative sign, according to the hysteresis direction. The negative areas compensate for the parts of positive trapezoids that extend beyond the boundaries of the polygon. Provided that the polygon is shifted into the first quadrant by some offsets h_{os} and b_{os} , the power loss in W m^{-3} caused by magnetic hysteresis effects can be computed with the frequency f , $M = 1024$, and circular padding by

$$\hat{p}_{\text{hyst}} = f \cdot \frac{1}{2} \sum_{i=0}^{M-1} b_i (h_{i-1} - h_{i+1}). \quad (1)$$

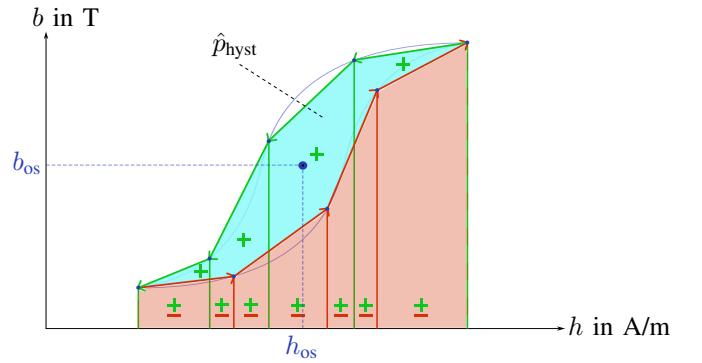


Fig. 1. Visualization of the shoelace formula applied to a bh polygon.

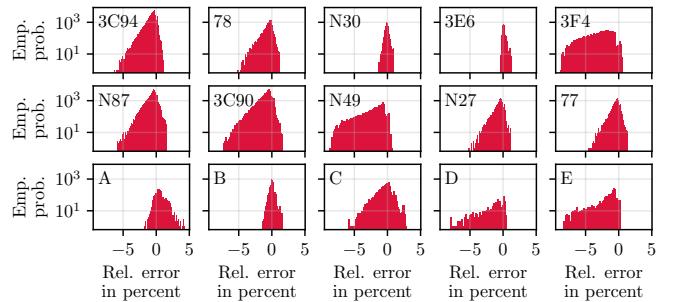


Fig. 2. Relative error $(\hat{p}_{\text{hyst}} - p)/p$ histogram between provided scalar p and \hat{p}_{hyst} calculated from the likewise provided bh polygon area.

When applying (1) on the given sequences $\mathbf{b}, \mathbf{h} \in \mathbb{R}^M$ with $M = 1024$, it becomes evident that the calculated area does not equal the provided loss measurements exactly. Fig. 2 shows the discrepancy with respect to the provided scalar loss p for all materials. The relative error ranges up to over 7 % for certain materials (e.g. 3F4, N49, D, E). Consequently, if merely an h -predicting model was to be identified, the lower bound on the rel. error would be significantly elevated by this circumstance alone.

Since the power losses calculated from neither the ground truth bh curve area (assuming ideal knowledge on the h sequence) nor the estimated area (\hat{h} reconstructed via a CNN) do perfectly match the provided loss measurement values (targets), an additional residual correction mechanism is added

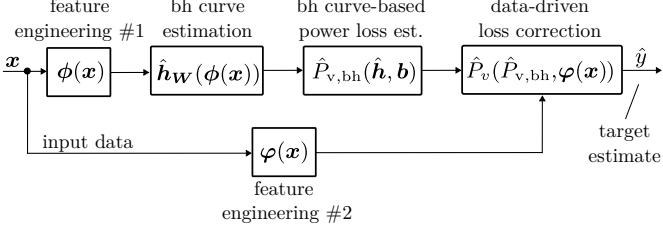


Fig. 3. Overview of the physics-inspired HARDCORE modeling toolchain.

to compensate for this. A high-level view on the proposed residual, physics-inspired modeling toolchain is depicted in Fig. 3, which is coined the HARDCORE approach (H-field and power loss estimation for Arbitrary waveforms with Residual, Dilated convolutional neural networks in ferrite COREs), and its details are discussed in the following.

II. MODEL DESCRIPTION

A residual CNN with physics-informed extensions is utilized for all materials. Such a CNN is trained for each material from scratch. Yet, the topology is unaltered across materials, signal waveforms, or other input data particularities.

A. One-dimensional CNNs for h -estimation

A 1D CNN is the fundamental building block in this contribution, which consists of multiple trainable kernels or filters per layer slided over the multi-dimensional input sequence in order to produce an activation on the following layer [2]. These activations denote the convolution (more precisely, the cross-correlation) between the learnable kernels and the previous layer's activation (or input sequence). In this stateless architecture, circular padding ensures that subsequent activation maps are of equal size. Circular padding can be utilized here instead of the common zero-padding as sequences denote complete periods of the b and h curve during steady state. Moreover, a kernel does not need to read strictly adjacent samples in a sequence at each point in time, but might use a dilated view, where samples with several samples in between are used. The dilated, temporal CNN update equation for the i -th filter's activation $a_i^{(l)}[k]$ at time k and layer l with the learnable coefficients $\mathbf{W}_i \in \mathbb{R}^{A \times \kappa}$ applied on A previous layer's filters, an uneven kernel size of $\kappa \in \{2x+1 : x \in \mathbb{N}_0\}$, and the dilation factor δ reads

$$a_i^{(l)}[k] = \sum_{p=0}^{A-1} \sum_{j=-(\kappa-1)/2}^{(\kappa-1)/2} \mathbf{W}_{i;(p,j)} \cdot a_p^{(l-1)}[k + j\delta]. \quad (2)$$

Since the task at hand does not require causality of CNN estimates along the time domain (losses are to be estimated from single b sequences), the sliding operation can be efficiently parallelized, and sequential processing happens merely along the CNN's depth. All 1D CNN layers are accompanied by weight normalization [3]. A conceptual representation of the 1D CNN for estimating \hat{h} is visualized in Fig. 6 (left part).

B. Feature engineering

The term feature engineering encompasses all preprocessing, normalization, and derivation of additional features in an observational data set. The input data contains the frequency f , the temperature T , the measured losses p as well as the 1024 sample points for the b and h waveforms. Especially the creation of new features that correlate as much as possible with the target variable (here, the h curve or the scalar power loss p) is an important part of most machine learning (ML) frameworks [4].

1) *Normalization:* As is typical in neural network training, all input and target features have to be normalized beforehand. All scalar and time series features are divided by their maximum absolute value that occurs in the material-specific data set, with the exception of the temperature and the frequency, which will be divided by 75°C and 150 kHz , respectively, regardless the material. Moreover, for an accurate h estimate, it was found to be of paramount importance to normalize each b and h curve again on a per-profile base in dependence not only on the ℓ_∞ norm of $|\mathbf{b}|$, but also on the maximum absolute b and h appearing in the entire material-specific data set. The latter two values are denoted b_{\lim} and h_{\lim} , and can be understood as material-specific scaling constants. In particular, the per-profile normalized b and h curves for a certain sample read

$$\mathbf{b}_n = \frac{\mathbf{b}}{\max_k |\mathbf{b}[k]|}, \quad \mathbf{h}_n = \frac{\mathbf{h}}{h_{\lim}} \cdot \frac{b_{\lim}}{\max_k |\mathbf{b}[k]|}, \quad (3)$$

with $h_{\lim} = \max_{i,k} |\mathbf{h}_i[k]|$, $b_{\lim} = \max_{i,k} |\mathbf{b}_i[k]|$, and i being the sample index in the entire material-specific data set. Then, \mathbf{b}_n is added to the set of input time series features, and \mathbf{h}_n is the target variable for the h estimation task.

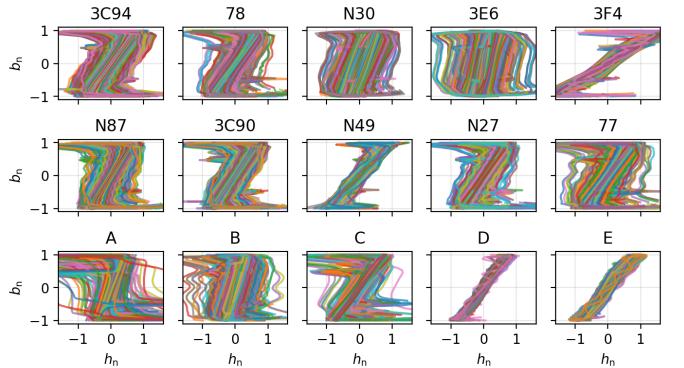


Fig. 4. Exemplary samples of the normalized b and h curves.

The \mathbf{b}_n over \mathbf{h}_n curves are displayed in Fig. 4, which underlines how the polygon area becomes roughly unified (no large area difference between samples). In the following, all features that get in touch with the model are normalized values without any further notational indication.

2) *Time series features (feature engineering #1):* As discussed in Sec. II-A, 1D CNNs build the core of the implemented model. The inputs to the CNNs are the (per-profile) normalized magnetic flux density b_n and the corresponding first and second order derivatives (\dot{b}_n and \ddot{b}_n) as time series. In a macroscopic measurement circuit context, \dot{b} corresponds to

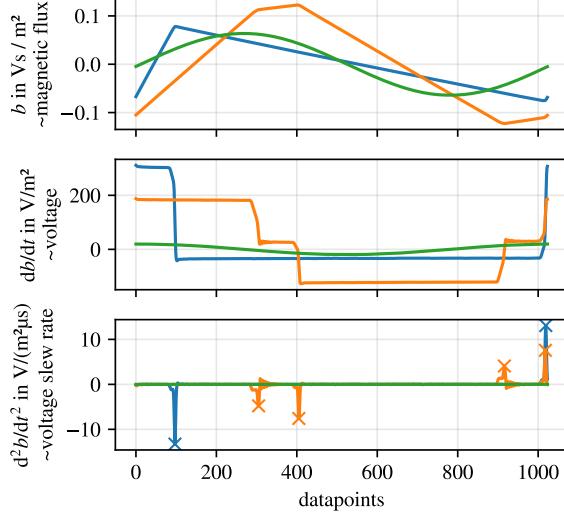


Fig. 5. Magnetic flux density examples and their first and second order derivatives for a sinusoidal, triangular and one unclassified waveform with a circuit-based interpretation in terms of their proportionality to magnetic flux, voltage and the voltage slew rate.

the applied magnetizing voltage throughout the measurement process of the data. Accordingly, \ddot{b} represents the voltage slew rate during the commutation of the switches in the test setup. Consequently, the second derivative allows to detect switching events and to characterize them according to their maximum slew rate. Fig. 5 shows, that the sinusoidal waveform (green) is generated without any fast transient switching behaviour, probably with a linear signal source. The nonsinusoidal examples show typical switching behaviour with different voltage slew rates during the single transitions and voltage overshoots as well as ringing. The second derivative of b informs the ML model about switching transition events and how fast changes in time are.

3) *Scalar features (feature engineering #2):* Although sequence-based CNNs take up the main share of the ML model size, scalar environmental variables also have a considerable impact on \mathbf{h} and p . While the temperature T is passed to the model unaltered (but normalized), the frequency is presented by its logarithm $\ln(f)$. The sample time $1/f$ is passed directly to the model. Furthermore, some b -derived scalar features are also passed to the model to feed in a priori knowledge. For example, the peak-to-peak magnetic flux Δb as well as the mean absolute time derivative $|\dot{b}|$ are directly fed into the network. Each waveform is automatically classified into "sine", "triangular", "trapezoidal", and "other" by consulting the form and crest factors, as well as some Fourier coefficients. The waveform classification is presented to the model by one hot encoding (OHE). A summary of all expert-driven input features is presented in Tab. I.

C. Residual correction and overall topology

The model topology comprises multiple branches that end in the scalar power loss estimate \hat{p} . An overview is sketched in Fig. 6. Two main branches can be identified: an h -predictor

TABLE I
UTILIZED INPUT FEATURES.

Time series features	Scalar features
mag. flux density b	temperature T
per-profile norm. b_n	sample time $1/f$
1st derivative \dot{b}_n	log-frequency $\ln(f)$
2nd derivative \ddot{b}_n	peak2peak Δb
$\tan(0.9 \cdot \tan(b_n))$	log peak2peak $\ln(\Delta b)$
	mean abs dbdt $ \dot{b} $
	log mean abs dbdt $\ln(\dot{b})$
	waveform (OHE)

and a wrapping p -predictor. The h -predictor utilizes both time series and scalar features with CNNs and multilayer perceptrons (MLPs), and estimates the full \mathbf{h} sequence. The p -predictor predicts p , on the other hand, and leverages the predicted magnetic field strength $\hat{\mathbf{h}}$ with the shoelace formula and a scaling factor. The latter accounts for the losses inexplicable by the bh -curve (recall Fig. 2), and is predicted by a MLP that utilizes the scalar feature set only.

The h -predictor merges time series and scalar feature information by the broadcasted addition of its MLP output to a part of the first CNN layer output. This effectively considers the MLP-transformed scalar features as bias term to the time-series-based CNN structure.

On the merged feature set, two further 1D CNN layers follow that end the transformation in a 1024-element sequence. The per-profile scaled b_n sequence from the set of input time series is element-wise added to this newly obtained estimation (residual connection). This results in the CNN model to merely learn the difference between h_n and b_n [5]. Eventually, this sequence becomes the h estimation \hat{h}_n when the sequence's average along the time domain, that is, across all 1024 elements, is subtracted from each element. This is a physics-informed intervention in order to ensure a bias-free h estimate \hat{h} after denormalization. Note that all such operations are still end-to-end differentiable with an automatic differentiation framework such as PyTorch [6], [7].

Since the resulting \hat{h} can only be trained to be as close as possible to the provided h sequence, which is not leading to the correct p ground truth (cf. Fig. 2), another MLP is branched off the scalar input feature set, and denotes the start of the p -predictor. This MLP inherits two hidden layers and concludes with a single output neuron. This neuron's activation, however, is not \hat{p} but rather an area scaling factor $s \in [-1, 1]$ to be embedded in the shoelace formula (1) with

$$\hat{p} = f \cdot (0.5 + (0.1 \cdot s)) \sum_{i=0}^{M-1} b_i (\hat{h}_{i-1} - \hat{h}_{i+1}). \quad (4)$$

Consequently, the p -predictor branch can alter the shoelace formula result by up to 10 % in positive and negative direction.

The p -predictor can be justified physically, when referring to Fig. 2 again. As the comparison shows, the hysteresis loss represents the total loss within a variation of -10 % to +5 %. The positive deviation ($\hat{p}_{\text{hyst}} > p$) indicates some measurement discrepancy between the measured loss and the given b and h curves. For parts of the negative deviations, a physical explanation can be found in eddy current losses, related to

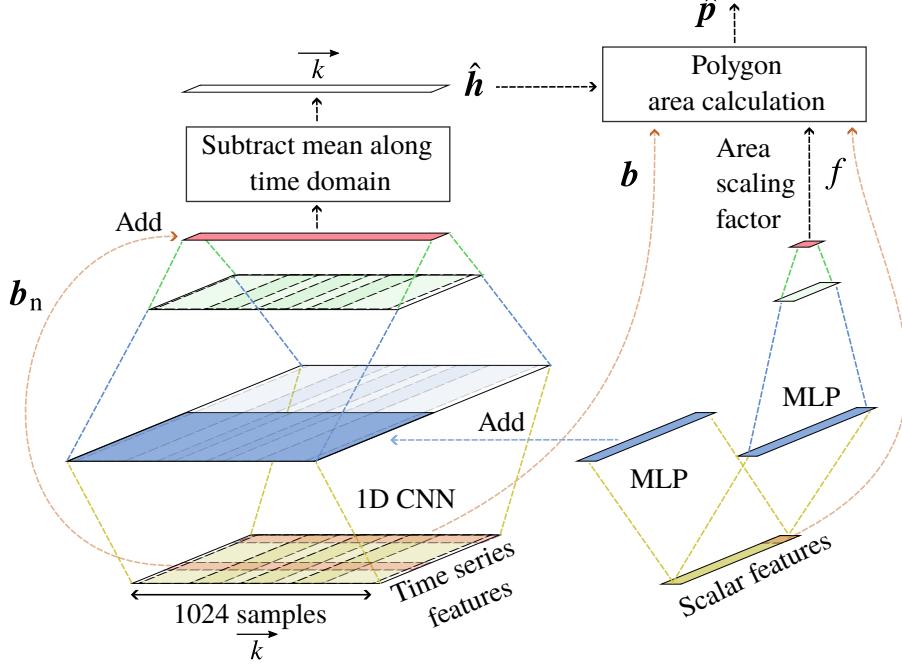


Fig. 6. The residual 1D CNN topology is shown while applied on time series and scalar features, which also contain engineered features from Tab. I.

a high dielectric constant and non-zero conductivity. Due to the small thickness of the used toroidal cores and the limited excitation frequency, eddy current losses are assumed to be of minor effect.

The physical interpretability of the intermediate estimate \hat{h} is a key advantage of the HARDCORE approach: First, it enables utilizing full h time series simulation frameworks (e.g., time domain FEM solvers). Secondly, for future designs of magnetic components with arbitrary shapes it becomes indispensable to accurately take into account also geometrical parameters of the core. This is only possible by distinguishing between the magnetic hysteresis and the (di)-electric losses.

D. Training cost functions

The training process involves two cost functions for a training data set with size N : First, the h estimation accuracy, which is assessed with the mean squared error (MSE) as

$$\mathcal{L}_{\text{MSE},h} = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} (\hat{h}_{i,n} - h_{i,n})^2. \quad (5)$$

Second, the power loss estimation accuracy is to be gauged. Despite the relative error being the competition's evaluation metric, the mean squared logarithmic error (MSLE) is selected

$$\mathcal{L}_{\text{MSLE},p} = \frac{1}{N} \sum_{n=0}^{N-1} (\ln \hat{p}_n - \ln p_n)^2 \quad (6)$$

in order to not overemphasize samples with a relatively low power loss [8]. As $\mathcal{L}_{\text{MSLE},p}$ also depends on \hat{h} through (4), the question arises, how both cost functions are to be weighted. In this contribution, a scheduled weighting is applied with

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{MSLE},p} + (1 - \alpha) \mathcal{L}_{\text{MSE},h}, \quad (7)$$

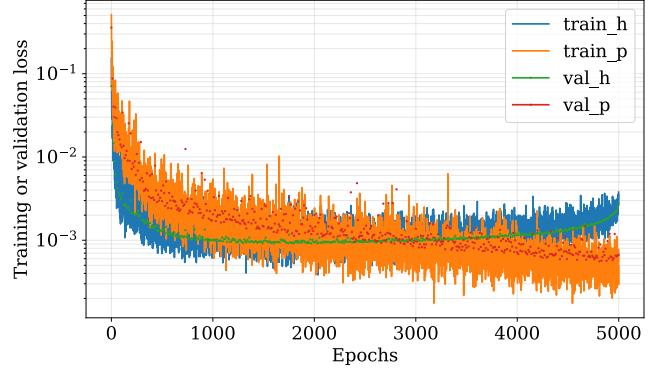


Fig. 7. Exemplary training and validation loss curve for material A, seed 0 and fold 3.

where $\alpha = (\beta \cdot i_{\text{epoch}})/K_{\text{epoch}}$ with $\beta \in [0, 1]$ denoting a hyperparameter scaling factor, K_{epoch} being the number of training epochs, and $i_{\text{epoch}} \in \{0, 1, \dots, K_{\text{epoch}} - 1\}$ representing the current epoch index. The scheduled weighting ensures that the model focuses on \hat{h} in the beginning of the training, where more information is available. Later though, the model shall draw most of its attention to the power loss estimate, possibly at the expense of the h estimation accuracy. A training example for material A is depicted in Fig. 7.

III. HYPERPARAMETERS, PARETO FRONT AND RESULTS

The proposed topology features several degrees of freedom in form of hyperparameters. An important aspect is the model size, which is defined by the number of hidden layers and neurons in each layer. A simple trial-and-error investigation

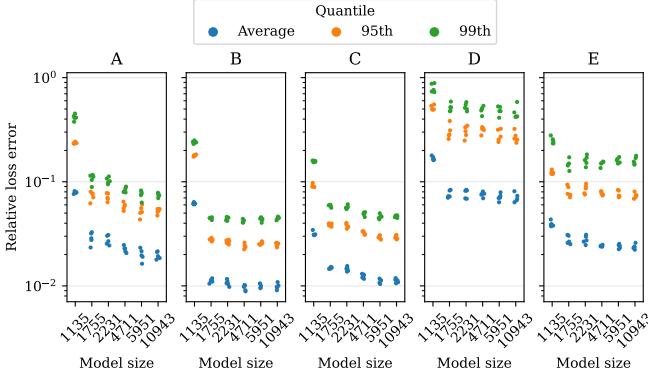


Fig. 8. Pareto front for the evaluation materials (A, B, C, D and E) showing model size (amount of parameters) vs. relative error of the power loss estimation.

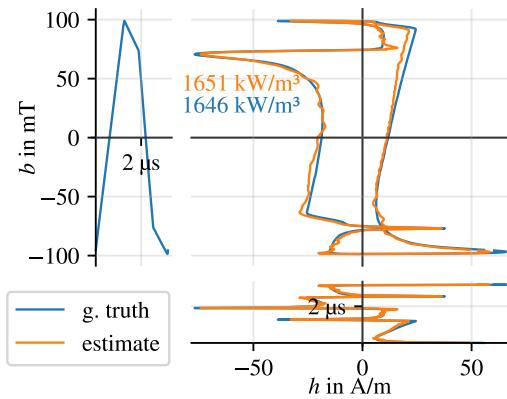


Fig. 9. Exemplary ground truth vs. estimated bh curve comparison.

can provide fast insights into the performance degradation that comes with fewer model parameters. In Fig. 8, several particularly selected model topologies are illustrated against their achieved relative error versus the inherited model size. The scatter in each quantile is due to different random number generator seeds and folds during a stratified 4-fold cross-validation. Topology variations are denoted by the amount of neurons in certain hidden layers. In addition, the largest topology has an increased kernel size with $\kappa = 17$, and the smallest topology has the second CNN hidden layer removed entirely (the green layer in Fig. 6).

A slight degradation gradient is evident as of 5 k parameters for materials A and C, whereas for the other materials the trend is visible only when removing the second hidden layer. Overall, the material performance scales strictly with the amount of training data available. Since fewer model parameters are a critical aspect, the chosen final model has 1755 parameters, which is at an optimal trade-off point on the Pareto front.

In Tab. II, the model size of a corresponding PyTorch model file dumped to disk as just-in-time (jit) compilation is reported. The exemplary bh -curve and h -curve estimation is shown in Fig. 9. Reported error rates come from the best seed out of five during a four-fold cross-validation ($\beta = 1$, $K_{\text{epochs}} = 5000$, Nesterov Adam optimizer). It shows effectively that any ma-

TABLE II
FINAL MODEL DELIVERY OVERVIEW

Material	Parameters	Training data	Model size	Relative error	
				Average	95-th quantile
A	1755	2432	43.13 kB	2.34 %	6.20 %
B	1755	7400	43.13 kB	1.10 %	2.68 %
C	1755	5357	43.13 kB	1.46 %	3.70 %
D	1755	580	43.13 kB	7.03 %	25.76 %
E	1755	2013	43.13 kB	2.51 %	7.10 %

terial can be modeled with the same topology at high accuracy as long as a critical training data set size is available (which is not the case for material D, see available training data in Tab. II). The final model is already a trade-off between model size and accuracy, such that in case one of the two criteria can be softened, the other can be further improved.

The final model delivery is trained on all training data samples (no repetitions with different seeds), and with $K_{\text{epoch}} = 10000$, $\delta = 4$, $\kappa = 9$. This final topology features a CNN with 12(TanH) → 8(TanH) → 1 (linear) kernels, a MLP with 11(TanH) neurons, and a p -predictor MLP with 8(TanH) → 1(TanH) neurons.

IV. CONCLUSION

A material-agnostic CNN topology for efficient steady-state power loss estimation in ferrite cores is presented. Since the topology remains unaltered across materials and waveforms at a steadily high accuracy, the proposed model can be considered universally applicable to plenty of materials. As long as sufficient samples of a material are available (roughly, 2000), the relative error on the 95-th quantile remains below 8 %. Thus, the contributed method is proposed to become a standard way of training data-driven models for power magnetics.

REFERENCES

- [1] B. Braden, "The surveyor's area formula," *The College Mathematics Journal*, vol. 17, no. 4, pp. 326–337, 1986.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information and Processing Systems (NIPS)*, vol. 60, no. 6, 2012, pp. 84–90. [Online]. Available: <https://doi.org/10.1145/3065386>
- [3] T. Salimans and D. P. Kingma, "Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks," ArXiv e-prints: 1602.07868 [cs.LG], 2016. [Online]. Available: <http://arxiv.org/abs/1602.07868>
- [4] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, p. 78, 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2347736.2347755>
- [5] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [6] A. Paszke, S. Gross *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.
- [7] A. G. Baydin, B. A. Pearlmutter *et al.*, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, pp. 1–43, 2018. [Online]. Available: <https://arxiv.org/abs/1502.05767>
- [8] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *Journal of the Operational Research Society*, vol. 66, pp. 1352–1362, 2015.

MagNet Challenge 2023 - Politecnico di Torino Proposal

Students: Alessio Giuffrida, Nicolò Lombardo, Fabio Marmello, Simone Morra, Marco Pasquale
Supervisors: Luigi Solimene, Carlo Stefano Ragusa

Abstract—A methodology for calculating ferrite core losses is introduced within the context of the 2023 MagNet Challenge. The challenge aimed to explore alternative approaches for estimating losses, emphasizing the use of machine learning methods and generalized Steinmetz methods. Our team aimed to integrate both statistical and physical approaches by incorporating Neural Networks and the Composition Waveform Hypothesis. The final selection was made among numerous approaches based on its simplicity and practical functionality.

I. INTRODUCTION

CARLES Steinmetz proposed the Steinmetz formula ($P = kf^\alpha B^\beta$) in the late nineteenth century to calculate time-averaged core losses in soft magnetic materials. The Steinmetz equation represents a straightforward fitting formulation based on three material-depending parameters and two input variables: the magnetic induction peak value and the frequency of the excitation waveform. However, the framework in which the Steinmetz equation was developed is extremely different from the actual application area of magnetics for power electronics converters. Different magnetic materials, such as ferrites, amorphous and nanocrystalline materials, different excitation waveforms, such as triangular, trapezoidal, and PWM waveforms, and even higher switching frequencies require a more detailed approach for the description of the magnetic power losses. The initiative to develop a comprehensive model for predicting these critical losses in various electrical engineering applications gained traction on the centenary of Steinmetz's passing. From early June, our team explored suitable adaptations for modeling ferrite losses. We believed that the machine-learning aspect of the solution should complement a more physics-based approach to the challenge.

II. MODELING CHOICES

The developed approach for predicting losses in sinusoidal, triangular, and trapezoidal waveforms relies on triangular waveforms as the foundational element. Initially, the data is divided based on each waveform type, and neural networks are employed to compute triangular losses for different input conditions. Sinusoidal losses are derived by scaling the previous results through regression. Additionally, a portion of trapezoidal losses is determined using the Composite Waveform Hypothesis (CWH) [1], by using the predicted triangular losses as inputs. Finally, the remaining trapezoidal signals are examined using a distinct neural network separate from the

triangular one. The flow chart of Figure 1 summarizes the workflow of the proposed loss computation procedure.

The project has been entirely developed in Python 3.10 using open-source libraries and packages. Concerning the neural networks, we opted to use Keras [2]. Keras is an efficient tool for creating, training, and using neural networks in the Python environment, and it's well-suited for dealing with the complexities of analyzing different types of waveforms and operating conditions, proper of the MagNet Challenge tasks. Keras simplifies the whole process of working with neural networks, making it easier for us to design and implement our models.

III. LOSS COMPUTATION PROCEDURE

A. Waveform Detection

The model starts by dividing the input induction waveforms into four different types:

- 1) Triangular waveforms.
- 2) Sinusoidal waveforms.
- 3) Trapezoidal waveforms, without near-zero derivative segments.
- 4) Trapezoidal waveforms, with a near-zero derivative segment.

These classifications are established by evaluating the form and crest factors of the induction waveforms and grouping the outcomes accordingly. Sinusoidal and triangular signals are easily discernible since their form and crest values are constant. All the waveforms that do not respect the requirements of sinusoidal and triangular waveforms are considered trapezoidal, and their further sub-division involves assessing the derivative of the signal. The methodology is related to the application of the CWH, in which an equivalent frequency is defined as:

$$f_{\text{eq}} = \frac{\left| \frac{dB}{dt} \right|}{2B_{\text{pkpk}}} \quad (1)$$

If a segment of the trapezoidal signal exhibits an equivalent frequency lower than a predetermined threshold, the signal is placed in the last category. The obtained classification allows the implementation of the proposed framework for the computation of the power losses under different excitation waveforms.

B. Triangular Waveforms

After identifying the triangular signals, calculations are performed to determine key parameters (scalar inputs) for

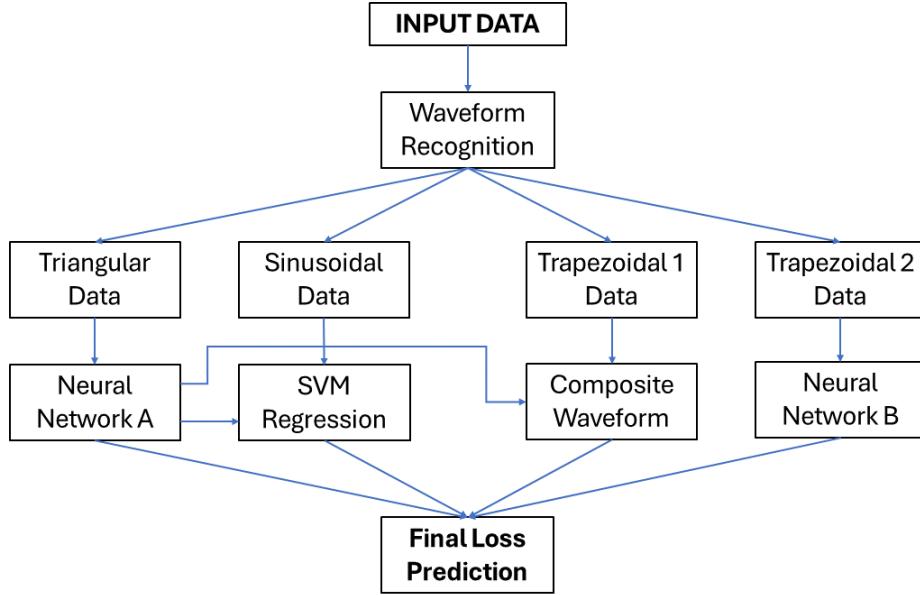


Fig. 1. Flowchart of loss prediction algorithm

predicting losses using the neural network. This involves computing the peak-to-peak induction value and duty cycle for each waveform. Subsequently, a data frame is structured to link each triangular waveform with its peak-to-peak induction, frequency, temperature, and duty cycle (B_{pkpk} , f , T , and δ). This data frame is then used as input for the neural network. The neural network output is the energy loss per cycle. Then, the power loss for each waveform is obtained by multiplying the correspondent energy loss by the frequency. Since the dependence of the energy loss on the induction and the frequency follows a power law, the logarithm operation is applied to these inputs and to the output. Regarding the training phase, performed on the training database of the five final materials, the data are shuffled and divided into training, validation, and testing sets. Then, the inputs are standardized to obtain compatible dimensions for all the different quantities. For all the sets, the mean and standard deviation of the training dataset are used for the standardization. The neural network structure has been preliminary determined by using a random search optimization, defining a Sequential Network in Keras described by Table I.

TABLE I
NEURAL NETWORK STRUCTURE FOR TRIANGULAR WAVEFORMS

	Number of neurons	Activation function
Layer 1	10	Tanh
Layer 2	16	Tanh
Layer 3	1	Relu

Every neural network for each material has 243 trainable parameters.

Once the network structure is defined and the other hyperparameters are set, the training of a neural network for each material is performed, with the Adam algorithm for the optimization of weights and biases, in order to minimize the

mean squared error on the validation datasets. Then, the neural network output is verified on the testing dataset.

C. Sinusoidal Waveforms

The detected sinusoidal waveforms underwent analysis with a Support Vector Regression (SVR) model. In this case, SVR involves linking the computation of losses from triangular to sinusoidal waveforms. Specifically, an examination was conducted to understand the relationship between symmetric triangular waveforms and sinusoidal waveforms with equivalent peak-to-peak induction, frequency, and temperature values. In fact, it is well-known that the dynamic magnetic loss terms are dependent on the time derivative of the induction ($\frac{dB}{dt}$) [4], [5]. The loss dependence on ($\frac{dB}{dt}$) is influenced by the main loss mechanism involved at a given frequency, peak-to-peak induction, and temperature for a given material. Theoretically, by assuming an energy loss dependence on the square of the time derivative of the induction ($W \propto (\frac{dB}{dt})^2$), the ratio between sinusoidal and symmetric triangular waveform losses is identified in $\pi^2/8$. However, this ratio is strongly influenced by the operating conditions, determined by the induction, frequency, and operating temperature, and is also dependent on the specific material properties. Figures 2 and 3 give an overview of the behavior of this ratio for the training set of Material B.

Since the dependence of this ratio on the input parameters is not easily determinable theoretically with the available data, our approach is based on modeling the sinusoidal-triangular loss ratio with an SVR. Thus, for a sinusoidal waveform at a given peak-to-peak induction, frequency, and temperature, we compute the correspondent energy loss of a symmetric triangular waveform. Then, we estimate the sinusoidal-triangular ratio by means of the trained SVR model, and finally, we combine the two outputs to obtain the energy losses of the sinusoidal waveforms.

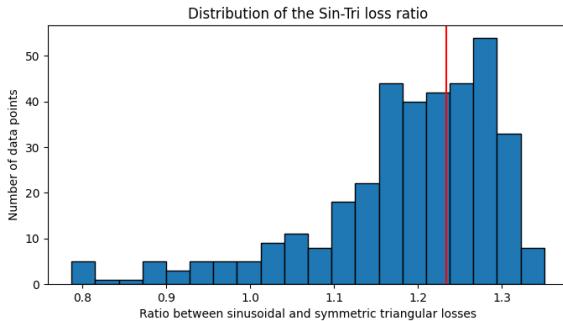


Fig. 2. Distribution of the sinusoidal-symmetric triangular loss ratio for the training set of Material B. The red line represents the theoretical value of $\pi^2/8$

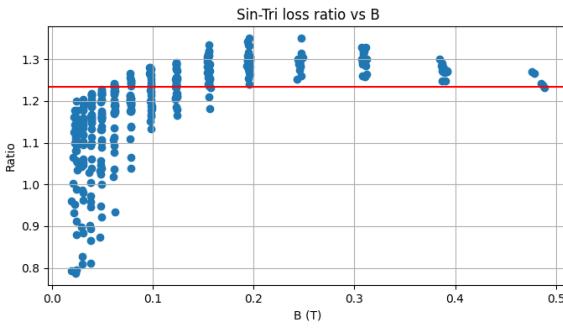


Fig. 3. Behavior of the sinusoidal-symmetric triangular loss ratio as a function of the peak-to-peak induction for the training set of Material B. The red line represents the theoretical value of $\pi^2/8$

For the SVR implementation, the Scikit-learn library has been used [3]. The Radial Basis Function (RBF) kernel is selected to deal with the non-linear nature of the regression problem. The inputs of the regression model are the peak-to-peak induction, frequency, and temperature. Even in this case, the training data are shuffled and divided into training, validation, and testing sets. Then, the inputs are standardized. An SVR model was trained for each material by computing the sinusoidal-symmetric triangular ratio on the training dataset under the same input conditions and using it as the target of the regression model.

The number of parameters required for defining the support vectors for each material model are reported in Table II.

TABLE II
NUMBER OF PARAMETERS OF THE SVR FOR EACH MATERIAL

	A	B	C	D	E
N. parameters	138	288	276	228	138

D. Trapezoidal Waveforms

In this case, trapezoidal waveforms were additionally differentiated depending on the presence of a segment with zero derivative. When examined with CWH, these segments display an equivalent frequency of zero, leading to non-physical losses. Hence, a distinct solution is required for this specific type of waveform.

1) *Without near-zero derivative segments:* These signal types were addressed by employing CWH and scaling the outcomes derived from triangular waveforms. The equations and procedures utilized align with those detailed in [1]. Notably, the distinction lies in the continuous execution of CWH across samples instead of the linearization of discrete segments. While this minimizes relative errors, it necessitates a longer computation time. Thus, the equivalent frequency is computed for each waveform sample, and for each, an energy loss value of a correspondent symmetric triangular waveform is computed by the previously trained neural network (under the same induction and temperature of the trapezoidal waveform). Then, the energy loss of the trapezoidal waveform is computed from the contributions of all the samples as

$$W = f \sum_i^{n_{\text{samples}}} W_i(B_{\text{pkpk}}, f_{\text{eq},i}, \text{Temp}) \Delta t_i. \quad (2)$$

A possible weakness of the CWH is that the computed equivalent frequency can exceed the frequency bounds of the training dataset. In this case, our approach consists of considering the energy losses independent of the frequency under the lower bound of the dataset since, typically, for the soft ferrites, the hysteresis loss term is still dominant up to some tens of kHz. On the other hand, for equivalent frequencies that exceed the upper bound of the dataset, we still take into account the output of the neural network, even if not trained in this frequency range, since it seems to be coherent with the typical physical behavior of the energy losses of ferrite materials below the MHz range. Despite the described limitations and the worst behavior in the context of our loss prediction procedure, the strength of the loss prediction for non-zero derivative trapezoidal waveforms is the adoption of the same neural network of the triangular waveforms and the application of the CWH, without requiring further training phase and further trained parameters.

2) *With a near-zero derivative segment:* A different approach was taken when identifying trapezoidal induction waveforms with a near-zero derivative. As CWH proved incapable of estimating losses within these segments, an additional neural network was constructed to tackle this issue. The input parameters for this neural network include the peak-to-peak induction, frequency, temperature, and the RMS value of the induction time derivative (B_{pkpk} , f , T , and $(\frac{dB}{dt})_{\text{RMS}}$). The procedure followed for the definition of the network structure is the same as adopted for the triangular networks. The obtained Sequential Network in Keras is described by Table III.

TABLE III
NEURAL NETWORK STRUCTURE FOR TRAPEZOIDAL WAVEFORMS WITH NEAR-ZERO DERIVATIVE SEGMENTS

	Number of neurons	Activation function
Layer 1	12	Tanh
Layer 2	12	Tanh
Layer 3	1	Relu

Every neural network for each material has 229 trainable parameters.

As for the networks for the triangular waveforms, the logarithm operation is applied to the peak-to-peak induction, the frequency, and the energy losses. The training data are shuffled and divided into training, validation, and testing sets. Then, the inputs are standardized to obtain compatible dimensions for all the different quantities. For all the sets, the mean and standard deviation of the training dataset are used for the standardization. The Adam algorithm is adopted for the optimization of weights and biases in order to minimize the mean squared error on the validation datasets. Then, the neural network output is verified on the testing dataset.

A remark is required for the trapezoidal induction waveforms with a near-zero derivative of Material D. Since just eight samples are available for the training phase, the neural network for the trapezoidal waveforms for this material cannot be trained. Thus, the CWH is also applied to the trapezoidal waveforms with a near-zero derivative for Material D.

IV. TOTAL NUMBER OF PARAMETERS

Table IV reports the total number of parameters required for the estimation of the power losses for each material by considering all the models adopted for the different waveform types.

TABLE IV
NUMBER OF PARAMETERS REQUIRED FOR THE LOSS PREDICTION OF EACH MATERIAL

	A	B	C	D	E
N. parameters	610	760	748	700	610

V. RESULTS

This section reports a brief overview of the result obtained on the Training folder dataset of Material B, highlighting the results obtained for each type of waveform. It should be noted that this dataset also includes the data adopted for the training of the neural network for the triangular waveform, for the trapezoidal induction waveforms with a near-zero derivative, and of the SVR for the sinusoidal waveforms. Table V resumes the main error values obtained for the four waveform types, while Figures 4, 5, 6, and 7 reports the error distribution of the four cases.

	Mean (%)	95 perc (%)	Max (%)
Triangular	0.87	2.19	5.80
Sinusoidal	1.04	2.32	12.11
Trapezoidal CWH	22.6	36.42	40.02
Trapezoidal NN	1.32	3.48	21.02

TABLE V
MEAN VALUE, 95TH PERCENTILE, AND MAX VALUE OF THE RELATIVE ERROR ON THE LOSS PREDICTION FOR MATERIAL B (TRAINING SET)

VI. CONCLUSION

The report analyzes the solution that the PoliTO team has reached during these months of the MagNet Challenge. Various implementation possibilities have been considered,

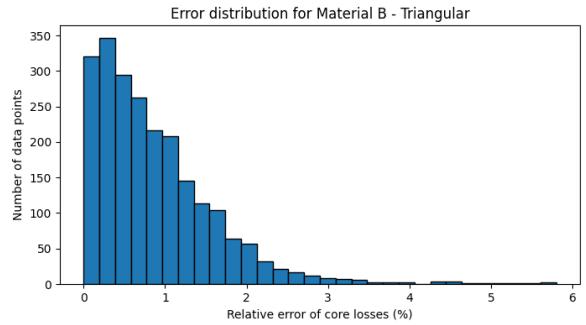


Fig. 4. Error distribution for the triangular waveforms (Material B).

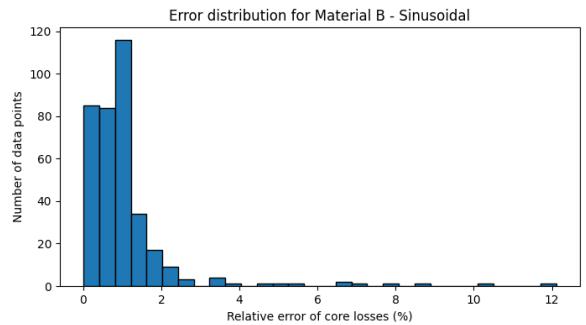


Fig. 5. Error distribution for the sinusoidal waveforms (Material B).

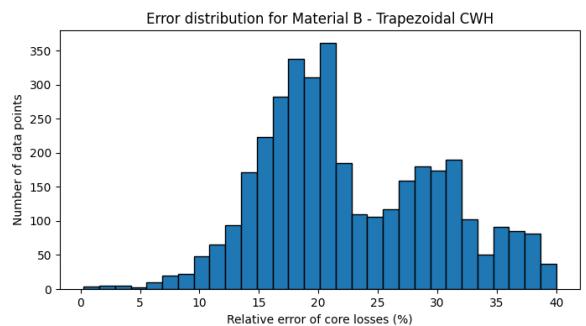


Fig. 6. Error distribution for the trapezoidal waveforms approached with the CWH (Material B).

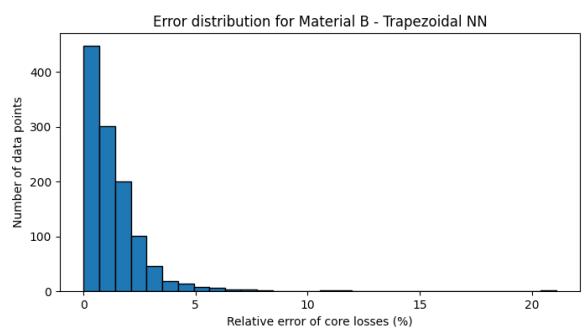


Fig. 7. Error distribution for the trapezoidal waveforms approached with the neural network (Material B).

aiming to strike the best compromise between result accuracy, computational time, and proposal simplicity. The identification

of losses in magnetic materials can be significantly improved based on the analysis conducted.

The adopted approach allows for a specific focus on each waveform type, as each presents specific aspects that make it challenging to identify a general method for all waveforms. The objective was to obtain differentiated loss prediction methods for each waveform, but with a limited number of input parameters to the neural networks. Where possible, efforts were directed towards the calculation of losses for a triangular waveform, as the neural network trained for triangular waveforms yielded satisfactory results.

The main challenge to address was the approach to trapezoidal waveforms, where two different paths were chosen depending on the presence of a near-zero derivative segment. Trapezoidal signals are those that exhibit a higher error in loss estimation, particularly for those without a near-zero derivative segment, where the CWH approach was adopted.

REFERENCES

- [1] T. Guillod, J. S. Lee, H. Li, S. Wang, M. Chen, and C. R. Sullivan, ‘Calculation of Ferrite Core Losses with Arbitrary Waveforms using the Composite Waveform Hypothesis’, in 2023 IEEE Applied Power Electronics Conference and Exposition (APEC), Mar. 2023, pp. 1586–1593. doi: 10.1109/APEC43580.2023.10131348.
- [2] ‘Keras: Deep Learning for humans’. Accessed: Dec. 30, 2023. [Online]. Available: <https://keras.io/>
- [3] ‘sklearn.svm.SVR — scikit-learn 1.3.2 documentation’. Accessed: Dec. 30, 2023. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>
- [4] H. Zhao, C. Ragusa, C. Appino, O. de la Barrière, Y. Wang, and F. Fiorillo, ‘Energy Losses in Soft Magnetic Materials Under Symmetric and Asymmetric Induction Waveforms’, *IEEE Transactions on Power Electronics*, vol. 34, no. 3, pp. 2655–2665, Mar. 2019, doi: 10.1109/TPEL.2018.2837657.
- [5] E. Barbisio, F. Fiorillo, and C. Ragusa, ‘Predicting loss in magnetic steels under arbitrary induction waveform and with minor hysteresis loops’, *IEEE Transactions on Magnetics*, vol. 40, no. 4, pp. 1810–1819, Jul. 2004, doi: 10.1109/TMAG.2004.830510.

MagNet Challenge Final Submission Protocol

Jacob Reynvaan, Martin Stoiber

Index Terms—Machine Learning, Magnetic loss, Neural Networks, MagNet, Steinmetz.

I. INTRODUCTION

THE MagNet Challenge, a practical initiative for engineering and science students, emphasizes the development of predictive models through machine learning. This challenge not only seeks to refine the accuracy of algorithms but also serves as a valuable platform for students to apply computational methods to real-world problems. In our participation, we initially faced the necessity to discard our original idea and shift our focus, adapting our strategy to the data. Our team's efforts have spanned various aspects of machine learning, from data analysis to model refinement. In our final presentation, we will share the insights we've gathered throughout this process.

II. DATA VISUALIZATION AND FEATURE EXTRACTION

The initial phase involved visualizing the dataset to understand its features and identify trends or irregularities. This was achieved by generating various plots to examine the data's distribution. After this analysis, attention was redirected towards developing feature extraction methods. The plotting of these features was then employed to evaluate their contribution to improving data separability, a critical factor in enhancing the subsequent training phase's efficiency and accuracy.

III. THE ORIGINAL IDEA

The initial proposal focused on utilizing Graph Neural Networks (GNN) in tandem with Symbolic Regression (SR) to accurately model magnetic core losses. (see figure 1) This approach involved training a GNN and applying SR to the edges of the graph, breaking down the complex problem into smaller, more manageable units. The concept was grounded in using SR, based on genetic algorithms, to find solutions through a novel integration with GNN.

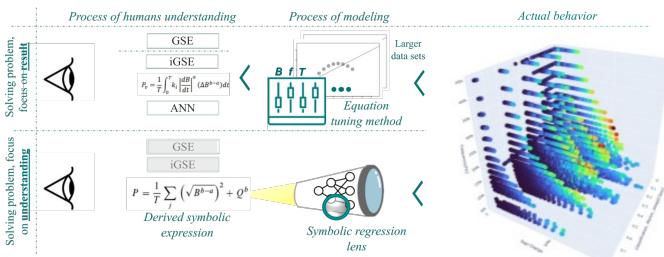


Fig. 1. From Data to Equation: A visual representation of the originally proposed method. The image depicts the transformation from raw data to a fully connected neural network (NN) model, and finally, the derived analytical expressions, general and easily perceived by humans.

A. Reasons Behind the Original Concept's Failure

The initial strategy entailed the utilization of GNN and SR as described in [1]. SR primarily relies on genetic algorithms, which do not guarantee an optimal solution. Hence, the proposed concept was to train a GNN and apply SR to the graph's edges, thereby breaking down the problem into smaller segments. However, due to the unsuitability of the given data for graph representation, we had to modify this approach. Figure 2 illustrates the arrangement of multiple waveforms, describing the magnetic flux, for each material grouped according to frequency and temperature. To determine the average core loss, polynomial regression was employed to obtain coefficients for these waveforms. Subsequently, SR was utilized for each coefficient to establish a formula that is dependent on both frequency and temperature. Unfortunately, the SR process did not yield any usable result. There are numerous factors contributing to the failure of this approach. The polynomial regression model was either underfitted or performed poorly. Additionally, it is likely that there does not exist a general description for each coefficient. Moreover, the coefficients may not solely depend on frequency and temperature.

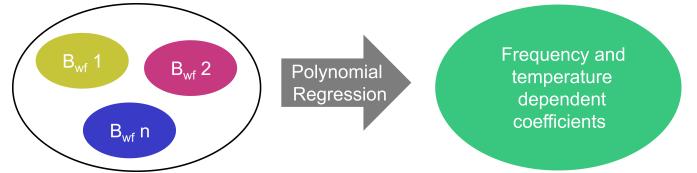


Fig. 2. The image portrays a process flow where multiple magnetic flux waveforms (B_{wf}), each with 1024 data points, are processed through polynomial regression to extract frequency and temperature-dependent coefficients, effectively simplifying the complex waveform data into a more manageable form for further analysis.

IV. DATA AUGMENTATION STRATEGIES FOR IMPROVED MODEL PERFORMANCE

A collection of 10 training materials was initially provided, followed by an additional set of 5 training materials, resulting in a total of 15 materials. Each material is assigned a unique material ID ranging from 0 to 14. The temperature is measured in degrees Celsius, while the frequency is measured in Hertz. Furthermore, the waveform of the magnetic flux is represented in Tesla and consists of 1024 data points. To minimize the impact of discontinuities, we employed linear interpolation to decrease the waveform from 1024 to a total of 256 data points. Following this, we performed Fast Fourier Transform (FFT) on the waveform sent in the respective frequency, resulting in 256 frequency points and 256 data points. These data points accurately reflect the energy levels at their corresponding

frequencies. A visual representation of this entire procedure is illustrated in figure 3.

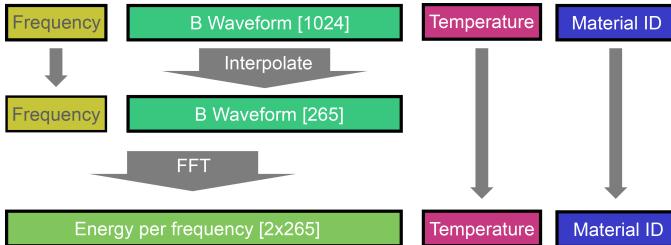


Fig. 3. The image illustrates the data preprocessing pipeline for a magnetic flux waveform analysis. Starting with the raw waveform at 1024 data points, it undergoes linear interpolation to reduce complexity, resulting in 256 points. A Fast Fourier Transform (FFT) is then applied, transforming the waveform into a frequency domain representation, which is subsequently normalized along with temperature and material ID for neural network training.

To effectively utilize this dataset in a neural network, it is necessary to normalize it. Prior to normalizing the entire dataset, each row is modified. The frequencies obtained from the FFT are divided by the respective maximum value, which is subsequently stored within the same row. Similarly, the energies derived from the FFT undergo the same normalization process. Henceforth the dataset has two more columns. Following this, the maximum energy, maximum frequency, temperature and material ID were normalized across all rows.

V. OPTIMIZATION OF NEURAL NETWORK ARCHITECTURE

In our pursuit to develop an effective neural network (NN) for modeling electronic components, we tailored the architecture by employing a fully connected neural network (FCNN). This approach has previously shown promise in the domain of electronic component modeling [2]. The initial network configuration, which featured an input layer with 2056 nodes, was systematically reduced through an iterative process, ultimately bringing the Input node count down to 516. Simultaneously, the corresponding hidden layers of the network were also methodically reduced in size, ultimately finding that a reduction in both size and depth significantly boosted performance. For the final Network size see Table II.

TABLE I
NEURAL NETWORK LAYER CONFIGURATION

Layer	Number of Neurons
Input Layer	516
Hidden Layer 1	512
Hidden Layer 2	128
Output Layer	1

These modifications, alongside the inherent benefits of FCNNs in capturing intricate data relationships, contributed to a refined model with reduced overfitting and enhanced generalization capabilities. The final, optimized architecture is depicted in figure 4. Structural adjustments to the network were complemented by experimentation with various activation functions, further tailoring the model's performance. After extensive testing, LeakyReLU emerged as the most effective choice, balancing non-linearity and computational efficiency, thereby contributing to the overall performance of the network.

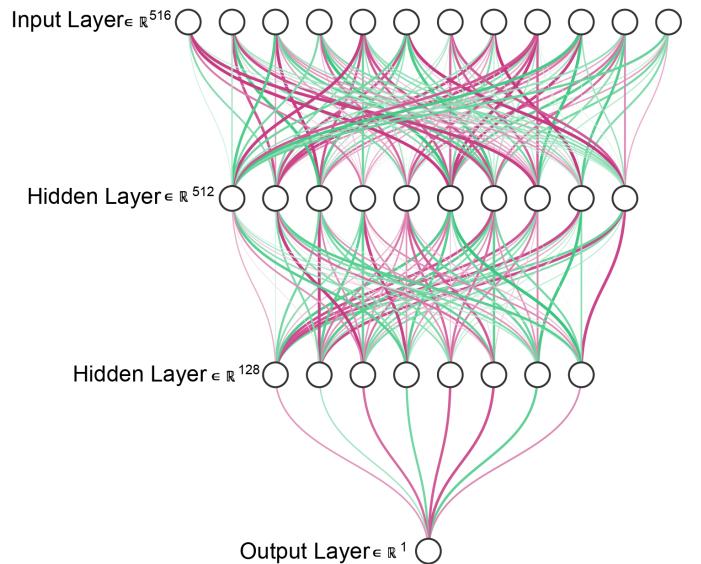


Fig. 4. The image visualizes the optimized neural network architecture with multiple input layers progressively reducing in size, from 516 nodes in the first layer to 512 in the second, and finally 128 nodes in the third layer, before converging to a single output node.

A. Refinement and Monitoring of NN Training

During the training phase, the model's progress was meticulously monitored using interactive visualization tools developed with Plotly. This approach was pivotal for real-time assessment and early detection of potential overfitting. The effectiveness of these tools is exemplified in the training and testing loss trends over time, as well as the adjustments to the learning rate, which are visually depicted in Figure 5. Our opti-

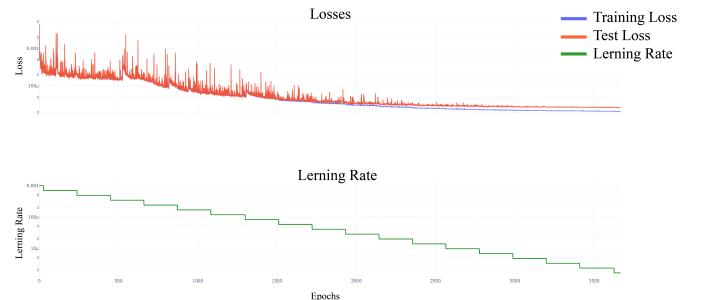


Fig. 5. The upper graph shows the training and test losses, both decreasing over time, indicative of the network's learning. The lower graph illustrates the learning rate schedule, which steps down at specific epochs, a technique often used to refine learning as the network's weights approach optimization. This combined visual representation aids in monitoring and adjusting the training process.

mization strategy was anchored around the AdamW optimizer, noted for its effectiveness with sparse gradients and inclusion of weight decay for regularization. In our implementation, the AdamW optimizer was set with an initial learning rate of 1×10^{-3} . In our final model configuration, weight decay was used, adhering to the default setting of 1×10^{-2} . Additionally, a learning rate scheduler was incorporated to optimize the training process further. The ReduceLROnPlateau scheduler from PyTorch's optimization suite was selected for its ability

to adjust the learning rate based on the model's performance. Specifically, the scheduler was configured with a cooldown period of 200 epochs and a learning rate reduction factor of 0.7. These combined strategies of real-time monitoring, advanced optimization techniques, and adaptive learning rate management significantly contributed to the effective and efficient training of our neural network model.

In the subsequent phase of the training process, a refining strategy was employed by reintroducing the previously trained model to a new training cycle, this time with an adjusted learning rate. Contrary to the initial higher learning rate used in the first phase, a slightly lower learning rate was used in this secondary phase. The aim was to fine-tune the model's weights and biases, allowing for more precise adjustments that could lead to enhanced performance. This approach is based on the principle that a lower learning rate in later stages of training can help the model converge more smoothly to a potentially better local minimum. An opportunity for the model to refine its learning in a more focused and detailed manner was provided by carefully reducing the learning rate for the already trained network. The outcome of this secondary training phase was encouraging. A discernible improvement in the model's performance metrics was observed, affirming the efficacy of gradually lowering the learning rate as part of the training regimen. This method underscores the importance of adaptive learning rate strategies in the iterative process of neural network optimization.

VI. COMPARATIVE ANALYSIS OF NEURAL NETWORK PREDICTION ACCURACY.

A. Accuracy Benchmarking on Original Data

Tabel II is segmented into four key metrics: average error, 95th percentile error, 99th percentile error, and maximum error. These metrics collectively offer a holistic view of the prediction accuracy and the distribution of errors within each dataset. The Average Error metric represents the mean prediction error for each dataset. The analysis revealed a spectrum of accuracy, with some datasets showing high precision, while others indicated lower predictive accuracy. The 95th Percentile Error provides insight into the upper range of the error distribution, indicating that 95% of the predictions fall below a certain error rate. This measure varied significantly across datasets, reflecting the differences in reliability. Further, the 99th Percentile Error extends our understanding to nearly the entire range of the data, revealing the error rate below which 99% of predictions fall. This metric is crucial as it highlights the presence of outliers or specific instances where prediction accuracy drastically decreases. Lastly, the Maximum Error observed in each dataset underscores the potential extent of inaccuracies, offering a perspective on the worst-case scenario in prediction accuracy for each dataset. The neural network's average prediction error is notably lower than the measurement error inherent in the data generation process. This implies that the predictions will consistently align within the accuracy range of the original measurements, effectively 'hitting' the correct values as per the measured data's accuracy. This highlights the model's precision in capturing the true nature

TABLE II
ERROR DISTRIBUTION ACROSS DATASETS

Material	Average	95th Percentile	99th Percentile	Maximum
3C90	0.50%	2.30%	5.96%	16.24%
3C94	0.29%	1.04%	2.27%	7.28%
3E6	0.23%	0.64%	1.70%	17.95%
3F4	0.75%	2.86%	10.18%	33.59%
77	0.53%	2.03%	6.11%	80.56%
78	0.52%	2.16%	5.67%	41.26%
N27	0.29%	0.97%	2.66%	10.89%
N30	0.36%	1.47%	3.10%	9.84%
N49	0.71%	3.00%	9.33%	39.32%
N87	0.49%	2.14%	4.77%	21.50%

of the datasets, despite the limitations of the measurement accuracy.

B. Adaptation and Prediction on Recent Datasets

The training of the neural network on newer datasets, despite initial challenges due to their smaller sizes, demonstrated considerable success. A diverse range of data, using both new and previously used datasets, was integrated, effectively enhancing the model's learning process. The results, as shown in Table III, illustrate the model's capability to accurately predict.

TABLE III
ERROR DISTRIBUTION FOR MATERIALS A-E

Material	Average	95th Percentile	99th Percentile	Maximum
Material A	0.83%	3.45%	6.76%	18.50%
Material B	0.38%	1.28%	3.09%	17.17%
Material C	1.01%	4.22%	11.07%	62.35%
Material D	1.44%	6.50%	20.74%	44.50%
Material E	0.29%	1.22%	3.61%	8.93%

VII. CONCLUSION

Our approach in the MagNet Challenge, from data visualization and feature extraction to meticulous neural network architecture optimization and training, has yielded a good predictive model. The model's performance, demonstrated by the error metrics across various datasets, underscores its robustness and accuracy. As the initial concept was discarded, strategies were adapted to the complexities of the data. The successful handling of both original and recent datasets underscores the adaptability and potential of our methods specifically for loss prediction in magnetic cores. It is noteworthy, however, that high-frequency materials seem to exhibit a slightly higher error margin. This observation suggests the need for further refinement in our model, particularly in its application to high-frequency materials.

ACKNOWLEDGMENT

We extend our thanks to the entire MagNet Team and the Jurors for their pivotal role in facilitating this remarkable challenge. We are grateful to SAL for providing the opportunity to contribute to the MagNet challenge. Special thanks to Milan Pajnic, Hélios Sanchis-Alepuz, Monika Stipsitz and Heinz Krenn for their invaluable discussions and encouragement that greatly enriched our experience.

REFERENCES

- [1] M. D. Cranmer, A. Sanchez-Gonzalez, P. W. Battaglia, R. Xu, K. Cranmer, D. N. Spergel, and S. Ho, "Discovering symbolic models from deep learning with inductive biases," *CoRR*, vol. abs/2006.11287, 2020. [Online]. Available: <https://arxiv.org/abs/2006.11287>
- [2] J. Reynvaan, M. Stipsitz, P. Skoff, T. Langbauer, and A. Connaughton, "A preliminary investigation into approximating power transistor switching behavior using a multilayer perceptron," in *2022 IEEE Design Methodologies Conference (DMC)*, 2022, pp. 1–5.

An Analytical Loss Model for Magnetic Cores Based on Vector Magnetic Circuit Theory

Chengbo Li, Wei Qin, Xiang Ma, Boyu Zhang, Zheng Wang, *Senior Member, IEEE* and Ming Cheng, *Fellow, IEEE*

Abstract- Currently, some empirical formula methods, such as the Steinmetz equation, have been widely used in calculating high-frequency magnetic core loss. Although these methods may provide satisfactory accuracy under some certain conditions, they cannot present clear physical interpretation and fail to account for the origin and variation of losses. In this article, we propose a new analytical model to predict core loss by employing a new magnetic circuit theory, i.e., vector magnetic circuit theory. In addition to reluctance, two new components, i.e., magductance and hysteresitance are proposed in the vector magnetic circuit, which are used to describe the eddy-current loss and the hysteresis loss, respectively. Thus, the clear physical significance is available with the proposed loss model. Given the geometrical and physical parameters of cores, the loss can be calculated mathematically. Furthermore, the proposed vector magnetic circuit based modeling method provides a different expression form for core loss, where the coefficients are with explicit physical meanings. Because only the measurement data are available with the competition, we employ the nonlinear fitting and Fourier decomposition to obtain the coefficients in the proposed new loss model. By comparison between the prediction results and the competition data, it has been verified that the proposed model could provide accurate and effective prediction of loss for high-frequency magnetic cores.

I. INTRODUCTION

Accurately predicting magnetic core losses is crucial to achieving optimal design of magnetic components [1]. Currently, there are three main methods used to calculate these losses: Steinmetz equation-based methods, loss separation model, and hysteresis model [2]. The loss separation model with the terms eddy current loss, hysteresis loss, and abnormal loss is primarily based on the G. Bertotti model [3]. On the other hand, the hysteresis model is based on the Preisach and Jiles-Atherton models [4-5], which are difficult to be applied in practical situations and lack universality [6].

For a long time, the empirical formula methods for core loss has been favored by engineers due to its practicality and simplicity [7-11], gradually moving from scientific research to the field of engineering applications. The emerging methods like neural networks are now becoming popular for their higher accuracy [12-14], which are significant in engineering applications. But they are sometimes lack of fundamental physical concepts in the calculation of magnetic core loss [15].

Previously, some researchers have tried to define more magnetic parameters in addition to reluctance, in order to model the characteristics of magnetic circuits such as losses more clearly [16-18]. However, almost all of them defined the magnetic parameters based on the duality between electrical circuit and magnetic circuit. Thus, those defined magnetic parameters are lack of the physical properties of magnetic circuit. For example, the “transference” parameter in the

Laithwaite magnetic circuit model [19] is solely mathematically dual to the inductance in the electrical circuit. Consequently, it is hard to use the “transference” concept to clearly describe some physical characteristics, e.g., power losses in magnetic circuit. Besides, the article of [19] also fails to characterize phase angle of magnetic flux in electromagnetic devices like transformers and electric machines.

Recently, Prof. Cheng's team from Southeast University proposed a new vector magnetic circuit theory [20-21], which merges the magnetic circuit and electromagnetic field theories, introducing a vector magnetic circuit model. Different from the previous scalar magnetic circuit with only reluctance, this new model comprises three components, i.e., reluctance, magductance, and hysteresitance. The model enables a deeper understanding of the characteristics and physical significance of magductance and hysteresitance components, as well as the magnetoelectric power law. The theory attempts to explain the origins of eddy current loss, hysteresis loss, and the phase shift between magnetomotive force and magnetic flux within the magnetic circuit. On this basis, this article first recalls the concept of vector magnetic circuit, then presents the analytical loss model for magnetic circuit, demonstrates the application process in the loss calculation of magnetic core. Finally, we verify these methods using competition data of materials 3E6 and N87.

II. VECTOR MAGNETIC CIRCUIT THEORY

Consistent with traditional scalar magnetic circuit theory, the reluctance, expressed by \mathcal{R} , represents the obstruction effect that magnetic fluxes encounter when they pass through a magnetic circuit. The formula of reluctance is as follows:

$$\mathcal{R} = \frac{l}{\mu S} \quad (1)$$

where μ is the magnetic permeability.

Based on the obstructing effect of a closed conductive coil on the alternating magnetic flux [20], the magductance has been defined as follows:

$$\mathcal{L} = \frac{\Gamma}{\Phi} = \frac{NQ}{\Phi} \quad (2)$$

where Γ is the charge linkage in N turns of closed conductive coil. N is the turns number of the coil. Q is the moving charge. Φ is the magnetic flux of magnetic circuit. The terminal characteristic of \mathcal{L} can be expressed as:

$$\mathcal{F}_L = \mathcal{L} \frac{d\Phi}{dt} \quad (3)$$

where \mathcal{F}_L is the magnetomotive force (MMF) of magductance component. The magductance could essentially represent the eddy-current effect in magnetic materials.

Based on the magductance component, we constructed the vector magnetic circuit theory and proposed the magnetoelectric power law in [20], which could be used to solve the long-standing problem to calculate the power of electromagnetic equipment through magnetic circuit directly. Recently, our team have further proposed the concept of hysteresis besides the magductance [21]. It could essentially represent the hysteresis effect in magnetic materials under alternating magnetic flux. The hysteresis can be defined as:

$$C = -\frac{\int \Phi dt}{\mathcal{F}_c} \quad (4)$$

where \mathcal{F}_c represents the MMF on the hysteresis component.

The unit of hysteresis C is $\text{Wb}\cdot\text{s}/\text{A}$, and its terminal characteristic is given as follows:

$$\mathcal{F}_c = -\frac{1}{C} \int \Phi dt \quad (5)$$

Fig. 1 shows the complete vector magnetic circuit proposed by our team, which comprises three fundamental passive components, namely reluctance, magductance, and hysteresis.

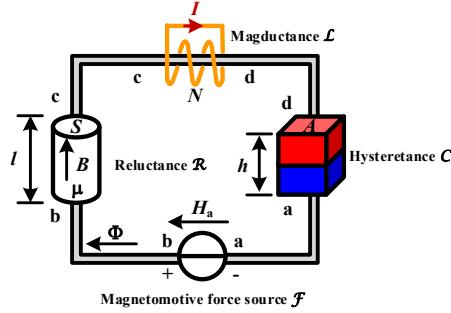


Fig. 1 Reluctance R , magductance L and hysteresis C are connected in series

III. ANALYTICAL LOSS MODEL FOR MAGNETIC CORE

We use the magnetic circuit to evaluate the power loss in magnetic core. When the magnetic circuit is excited by a steady-state sinusoidal magnetic electromotive force, the following steady-state expression can be obtained using the phasor method:

$$\dot{\mathcal{F}} = \mathcal{R}\dot{\Phi} + j\omega\mathcal{L}\dot{\Phi} + j\frac{1}{\omega C}\dot{\Phi} \quad (6)$$

where ω is the angular frequency.

According to the magnetoelectric power law [20], the active power of the magnetic circuit can be obtained:

$$P = P_L + P_C = \omega \left(X_L \|\dot{\Phi}\|^2 + X_C \|\dot{\Phi}\|^2 \right) = \omega^2 \frac{N^2}{R} \|\dot{\Phi}\|^2 + \omega \frac{h \sin \gamma}{\mu A} \|\dot{\Phi}\|^2 \quad (7)$$

where R is the resistance of the magductance component, γ is the hysteresis angle of the hysteresis component [21], h is the height of the hysteresis component, and A is the cross-sectional area of the hysteresis component. Eq. (7) provides a new analytical model to calculate the loss in magnetic circuit, consisting of two parts: one part is P_L on magductance component, corresponding to the eddy current loss. The other one is P_C on hysteresis component, corresponding to the hysteresis loss.

The reactive power of the magnetic circuit is:

$$Q = \omega \mathcal{R} \|\dot{\Phi}\|^2 \quad (8)$$

For the rectangular section shown in Fig. 2, we combine vector magnetic circuit theory and finite Fourier transform method [22]. Considering the spatial distribution of magnetic field strength and induced current density J , the white rectangular section in Fig. 2 is composed of different orange rectangular magnetic circuit components stacked together. These magnetic circuits are independent of each other. For each magnetic circuit element, the reluctance \mathcal{R}_{mn} in each circuit represents the obstruction effect on the magnetic circuit, the magductance \mathcal{L}_{mn} represents the eddy current effect in the magnetic circuit, and the hysteresis loss C_{mn} represents the hysteresis loss in the magnetic circuit.

The analytical expressions for parameters of each magnetic circuit component are given as:

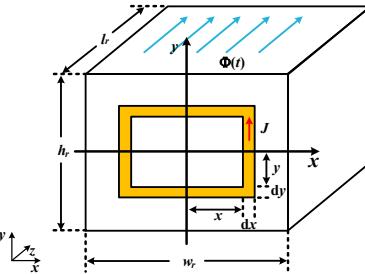


Fig. 2 Mathematical model of rectangular cross-section magnetic circuit

$$\mathcal{R}_{mn} = \frac{\pi^4 m^2 n^2 l_r}{64 \mu w_r h_r} \quad (9)$$

$$\mathcal{L}_{mn} = \frac{\pi^4 m^2 n^2 \sigma l_r}{64 w_r h_r \left(\frac{m^2 \pi^2}{w_r^2} + \frac{n^2 \pi^2}{h_r^2} \right)} \quad (10)$$

$$C_{mn} = \frac{64 \mu w_r h_r}{\cos \gamma \pi^4 m^2 n^2 l_r} \quad (11)$$

where w_r is the width of a rectangular cross-section, h_r is the height of the rectangular section, l_r is the length of the rectangular section, σ is the conductivity of the magnetic core, m is an odd number, and n is an integer.

Given the specific geometric shape, conductivity, and permeability of the magnetic circuit, the magnetic circuit parameters can be calculated using Eqs. (9)-(11). In consequence, the core loss can be calculated as:

$$P = \sum_{n=1,3,5,\dots}^{\infty} \sum_{m=1,3,5,\dots}^{\infty} (P_L + P_C) = \sum_{n=1,3,5,\dots}^{\infty} \sum_{m=1,3,5,\dots}^{\infty} (\omega^2 \mathcal{L}_{mn} \|\dot{\Phi}_{mn}\|^2 + \frac{\|\dot{\Phi}_{mn}\|^2}{C_{mn}}) \quad (12)$$

In reference [21], we validated the effectiveness of the analytical model using silicon steel sheets. Due to space limitations, this article will not provide more details.

IV. APPLICATION METHODS

Sections II and III presented the concept of vector magnetic circuit theory and analytical loss model. In practical applications, accurate calculation of magnetic circuit parameters is a new challenge. The literature [23] has identified various factors that influence magnetic core loss, including magnetic core material, size, magnetic density, frequency, and excitation type. However, the material manufacturers provide limited data, typically consisting of

fixed descriptions of loss curves and magnetic permeability change graphs with incomplete values over a wide range. Therefore, the loss calculations have been conducted by parameters fitting for magductance and hysteretance in Eq. (7), which is different from existing loss models.

A. The Impact of Frequency

The vector magnetic circuit theory can be employed in turn to derive the values of magductance and hysteretance, as follows:

$$P_v = k_e f^2 + k_h f \quad (13)$$

$$\chi_L = \frac{k_e f^2}{\omega \|\dot{\Phi}\|^2} = \frac{k_e f}{2\pi \|\dot{\Phi}\|^2} \quad (14)$$

$$\chi_C = \frac{k_h f}{\omega \|\dot{\Phi}\|^2} = \frac{k_h}{2\pi \|\dot{\Phi}\|^2} \quad (15)$$

where P_v represents the volumetric loss of the magnetic circuit, k_h is the hysteresis loss coefficient, k_e is the eddy current loss coefficient, and f is the frequency. Using Eqs. (7), (12)-(15), we have derived the loss frequency relationship curve for 3E6 material exposed to sinusoidal excitation with a magnetic density amplitude of 0.1T at 25°C, as depicted in Fig. 3. Similarly, Fig. 4 displays the loss frequency relationship curve for N87 material under sinusoidal excitation with a magnetic density amplitude of 0.035T at 25°C. The “Pred_Volumetric_Loss” is the predicted total power loss, the “Pred_Hysterisis_Loss” is the predicted hysteresis loss, the “Pred_Eddycurrent_Loss” is the predicted eddy-current loss, and the “Competition Volumetric Loss” is the total loss data provided by the competition organizer. The numerical values in the labels of the figures are arranged from left to right, which represent the predicted volume loss at each data point from low to high frequency, along with the percentage of error compared to the corresponding competition data.

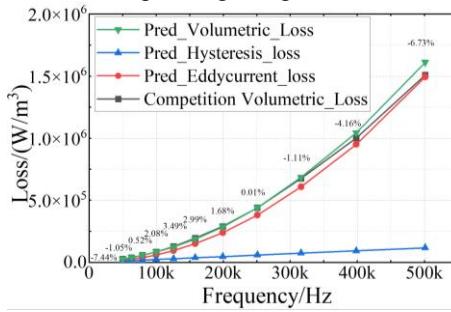


Fig. 3 The loss frequency relationship curve for 3E6 material exposed to sinusoidal excitation with a magnetic density amplitude of 0.1T at 25°C

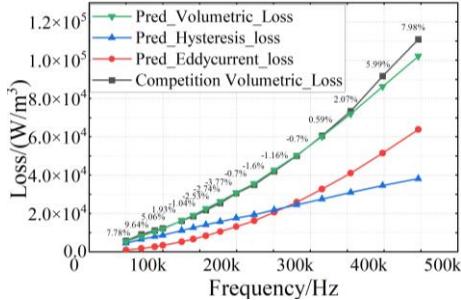


Fig. 4 The loss frequency relationship curve for N87 material exposed to sinusoidal excitation with a magnetic density amplitude of 0.035T at 25°C

The average values of errors shown in Figs. 3 and 4 are -0.89% and 2.88%, respectively. The effectiveness of our method within the wide frequency range under sinusoidal excitation has been proved. Furthermore, the method presented in this article allows for the determination of specific values of eddy current loss and hysteresis loss, as well as the patterns of frequency variation. For instance, comparing the eddy current losses in Figs. 3 and 4, we can see that the eddy current losses have always been the primary loss for 3E6 material. In contrast, for N87 material, as the frequency increases, the proportion of eddy current loss gradually exceeds hysteresis loss, highlighting that high conductivity magnetic core materials experience severe eddy current loss at high frequencies, while even low conductivity materials will experience more obvious eddy current loss at higher frequencies.

B. The Impact of Magnetic Density

Saturation has long been a significant source of nonlinearity in magnetic circuits [23]. When exposed to high frequency and intense magnetic fields, local saturation can occur within the magnetic core. Unfortunately, material manufacturers do not offer a quantitative assessment of the impact of this saturation, and only a limited amount of data can be obtained through experimental measurements. Our study addresses this issue by utilizing finite polynomials to fit changes in the eddy current coefficient and hysteresis coefficient with magnetic density amplitude. The specific form of our approach is outlined below:

$$k_e = a_1 B_m^3 + b_1 B_m^2 + c_1 B_m + d_1 \quad (16)$$

$$k_h = a_2 B_m^3 + b_2 B_m^2 + c_2 B_m + d_2 \quad (17)$$

where B_m is the amplitude of the sine wave excitation, and a_1 , a_2 , b_1 , b_2 , c_1 , c_2 , d_1 , d_2 are the fitting parameters. By utilizing Eqs. (16)-(17) in conjunction with Eqs. (13)-(15), we are able to derive the magnetic circuit parameters for various levels of magnetic flux density. This process enables us to create a correlation curve between loss, frequency, and magnetic density for N87 material under sine wave excitation at 25 °C, as shown in Fig. 5.

It should be mentioned that the aforementioned method can be used to determine the magnetic circuit parameters for different temperature states using Eqs. (13)-(17). Due to space limitations, we only show the case of 25 °C here.

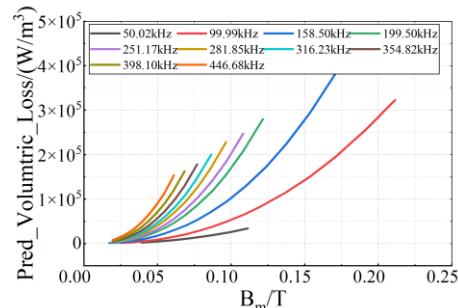


Fig. 5 The correlation curve between loss, frequency, and magnetic density for N87 material under sine wave excitation at a temperature of 25 °C

The average discrepancy between the predicted and experimental loss data on N87 material with sine wave excitation at 25 °C is 4.67%. Fig. 5 reveals that the loss also

increases gradually with an increase in magnetic density under the same frequency. Similarly, an increase in frequency also leads to a gradual increase in loss under the same magnetic density. If there are sufficient data available, a three-dimensional surface figure can be formed by combining the loss, frequency, and magnetic density.

C. The Impact of Excitation Type

In practical applications with high frequencies, triangular and trapezoidal wave excitations are commonly utilized. To address the inaccuracies in traditional Steinmetz mode calculations under these waveforms, some scholars have proposed the iGSE (improved general Steinmetz equations) [9] and i²GSE (improved-improved general Steinmetz equations) [11] models using empirical formula methods in recent years. Differently, this article proposes a new analytical loss modeling method by using vector magnetic circuit theory.

Previously, we obtained the magnetic circuit parameters under sinusoidal excitations at varying frequencies and magnetic densities. In signal processing, Fourier decomposition is often used to transform arbitrary waveform into a superposition of sine waves with different frequencies. Some scholars have attempted to use this method for magnetic circuit analysis before, but their results have not been satisfactory [8], [24-25]. From a vector magnetic circuit perspective, we believe that signals can be decomposed through Fourier transform, but magnetic circuit parameters cannot. Additionally, an arbitrary waveform cannot produce multiple magnetic permeabilities in a definite working state of the magnetic circuit. Therefore, the hysteresis coefficient k_h and eddy-current coefficient k_e under any excitation are primarily related to the peak value of the current excitation, which can be expressed as magnetic circuit parameters:

$$k_e = a_1 B_{peak}^3 + b_1 B_{peak}^2 + c_1 B_{peak} + d_1 \quad (18)$$

$$k_h = a_2 B_{peak}^3 + b_2 B_{peak}^2 + c_2 B_{peak} + d_2 \quad (19)$$

$$\chi_{L_n} = \frac{k_e f_n^2}{\omega_n \|\dot{\Phi}\|^2} = \frac{k_e f_n}{2\pi \|\dot{\Phi}\|^2} \quad (20)$$

$$\chi_C = \frac{k_h f_n}{\omega_n \|\dot{\Phi}\|^2} = \frac{k_h}{2\pi \|\dot{\Phi}\|^2} \quad (21)$$

where B_{peak} is the magnetic density peak of arbitrary waveform, and the quantities with n are the magnetic circuit parameters corresponding to the n th harmonic after Fourier decomposition.

Eqs. (20) and (21) demonstrate that when various sine waves with different amplitudes and frequencies are obtained through Fourier decomposition, the magnetic circuit will generate magductance corresponding to different frequencies. Eqs. (20) and (21) also show that the hysteresis of the magnetic circuit is only related to the excitation's B_{peak} value. In summary, the eddy-current loss can be superimposed while the hysteresis loss cannot be superimposed. This judgement agrees with the physical characteristics of hysteresis loss and eddy current loss [26]. On the other hand, this part cannot be analyzed using the existing empirical formula methods.

By combining Eqs. (20), (21) and (7), we can obtain that the loss corresponding to arbitrary waveform excitation can be expressed as follows:

$$P_v = \sum_1^n \omega_n (\chi_{L_n} + \chi_C) \|\dot{\Phi}\|^2 \quad (22)$$

As per the computation prescribed in Eq. (22), the error distribution histograms of 3E6 and N87 materials were obtained under arbitrary excitation waveforms at multiple magnetic-density points and multiple frequency points at a standard temperature of 25 °C, as shown in Figs. 6 and 7.

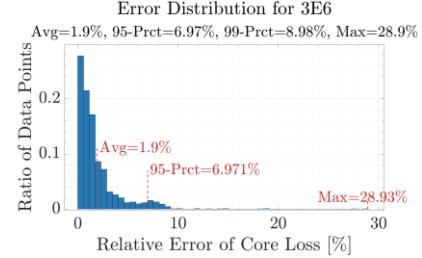


Fig. 6 The error distribution histograms of 3E6 materials were obtained under arbitrary waveform excitation at multiple magnetic density points and multiple frequency points at a standard temperature of 25 °C

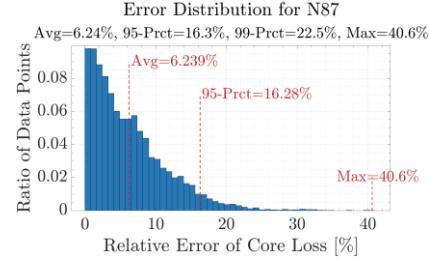


Fig. 7 The error distribution histograms of N87 materials were obtained under arbitrary waveform excitation at multiple magnetic density points and multiple frequency points at a standard temperature of 25 °C

It can be seen that the error of 95% of the data for material 3E6 is less than 6.972%, while the error for N87 material is 16.28% for the same percentage. After comparing Figs. 6 and 7, the average error, 95-prct error, and maximum error for N87 material are all higher than those of 3E6 material. Based on material properties, the most significant difference between these two materials is their conductivity, which varies by 100 times. Additionally, the comparison between frequency in Figs. 3 and 4 reveals a notable difference in the proportion of eddy current losses in the total losses of the two materials. This raises questions about the accuracy of the current algorithm in materials with low eddy current losses and the source of data with significant errors in the two materials. To answer these questions, we conducted further research.

Through our analysis of maximum error values, it was discovered that the N87 material experiences a loss four times greater under 80% duty cycle triangular wave excitation in comparison to 60% duty cycle triangular wave excitation. This difference is illustrated in Fig. 8, which depicts two types of triangular waves with equivalent frequency and magnetic density peak sizes. On the other hand, the 3E6 material only experiences a 20% difference under the same conditions.

The duty cycles of the two types of excitation waveforms are different, which are 60% and 80% respectively. But it is the difference in dB/dt values that results in the difference in hysteresis loss. As explained in [26], the hysteresis loss is caused by rotational friction of magnetic domains. This means that the faster the rotation speed of magnetic domain is, and

the greater the hysteresis loss is. As a result, the materials like N87 having a significant proportion of hysteresis loss, are more affected by dB/dt . Currently, this article is still in the stage of qualitatively analyzing the impact of dB/dt on hysteresis loss. Our next step is to study how we can quantitatively consider this impact with our method.

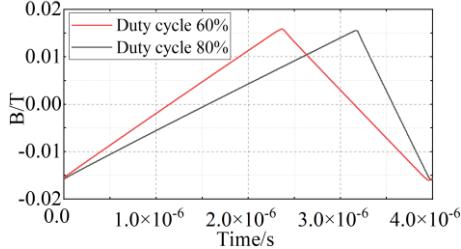


Fig. 8 Two types of triangular waves with equivalent frequency and magnetic density peak sizes.

V. COMPETITION CONTENTS

According to the methodology elucidated in this article, with regard to the final materials of A, B, C, D and E that were provided in the competition, we derived five distinct error distribution histograms from the training dataset. These histograms are graphically displayed in Fig. 9.

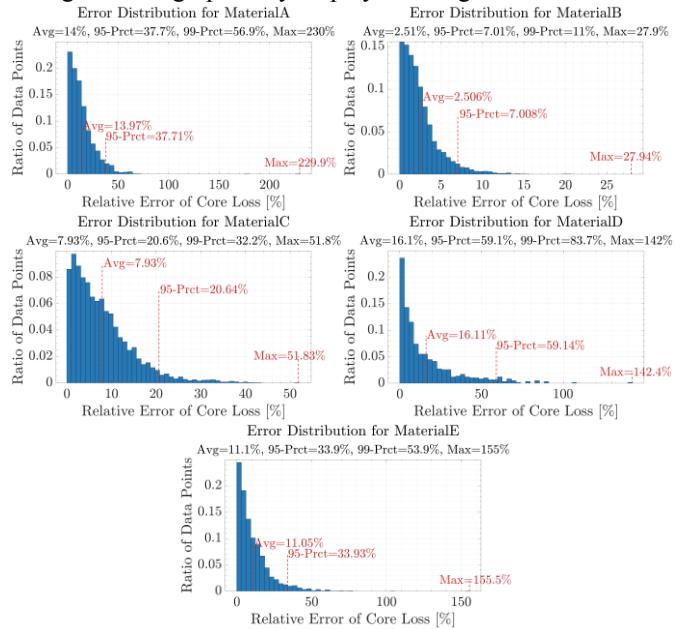


Fig. 9 The error distribution histograms of materials A, B, C, D and E from the training data.

The model size and numbers of parameters for each material (A, B, C, D, and E) are shown in Table I.

TABLE I

THE NUMBERS OF PARAMETERS FOR EACH OF THE MATERIAL

Material	A	B	C	D	E
Numbers of parameters	81	56	61	23	53
Model size (Byte)	5125	4421	4542	3255	4404

The statistical table of data points used for training and testing is shown in Table II.

TABLE II
THE AMOUNT OF TRAINING AND TESTING DATA

Material	A	B	C	D	E
Training data	2432 (101/69 4/1637)	7400 (364/225 3/4783)	5357 (215/167 9/3463)	580 (145/400/ 35)	2013 (57/667/ 1289)
Testing data	7651 (334/21 74/513)	3172 (147/980/ 2045)	5357 (212/175 4/3394)	7299 (61/2247/ 4991)	3738 (107/120 5/2426)

(The numbers in brackets denote the number of data points for sinusoidal waves, triangular waves, and trapezoidal waves, respectively.)

Based on our analysis in Fig. 9, the model is more effective for material B than material D. There are three sources of errors that we've identified:

1. The data used to train material B is more adequate than the data used to train material D. The results in more accurate curve fitting of eddy current loss and hysteresis loss with magnetic density and frequency.

2. The proportion of sine wave data used in the training is very small compared to other waveforms like triangular and trapezoidal waves. Although we can calculate the excitation of some other waveforms through Fourier decomposition, there is still a lack of quantitative research for excitation with large dB/dt .

3. For the final 5 materials, our model is relatively accurate in predicting materials with a high proportion of eddy current losses. Therefore, we can infer that material B has the highest conductivity among the 5 materials, while material D has a relatively small conductivity among the 5 materials.

VI. CONCLUSION

The vector magnetic circuit theory provides a different prospective and modeling method for predicting power loss in magnetic circuits. Different from existing core loss modeling methods, the proposed vector magnetic circuit based modeling method offers the following features: Firstly, it presents an analytical solution to evaluate the eddy-current loss and the hysteresis loss for magnetic cores by introducing the concepts of magductance and hysteretance in magnetic circuits. Given the complete geometrical and physical parameters of magnetic cores, the magductance and hysteretance parameters can be directly calculated, which can predict the eddy-current loss and hysteresis loss with magnetoelectric power law. Secondly, the proposed method offers the clear physical descriptions, i.e., magductance and hysteretance for eddy-current loss and hysteresis loss, which are not available by the existing loss modeling methods with abstract loss coefficients. Thirdly, the proposed method provides a new expression form for core loss. Even the material parameters are not available, the nonlinear fitting can be employed to evaluate the loss parameters and thus derive the loss model effectively.

Last but not least, we are very grateful to the organizers for organizing such a competition in 100-year honor of Prof. Charles P. Steinmetz (1865-1923). We hope that the vector magnetic circuit theory can provide an effective analytical modeling method for analyzing and calculating core losses.

REFERENCES

- [1] D. C. Jiles, "Frequency dependence of hysteresis curves in conducting magnetic materials," *J. Appl. Phys.*, vol. 76, pp. 5849–5855 1994.
- [2] A. Krings and J. Soulard, "Overview and comparison of iron loss models for electrical machines," *J. Elect. Eng.*, vol. 10, pp. 162–169, 2010.
- [3] G. Bertotti, "General properties of power loss in soft magnetic material," *IEEE Trans. Magn.*, vol. 24, no. 1, pp. 621–630, Jan. 1988.
- [4] D. C. Jiles and D. L. Atherton, "Theory of ferromagnetic hysteresis," *Magn.Magn. Mater.*, vol. 61, pp. 48–60, 1986.
- [5] A. Courtay, *The Preisach Model*. Austin, TX, USA: Analogy Inc., 1999.
- [6] S. Barg, K. Ammous, H. Mejibri, and A. Ammous, "An improved empirical formulation for magnetic core losses estimation under nonsinusoidal induction", *IEEE Trans. on Power Electron.*, vol. 32, no. 3, pp. 2146-2154, Mar. 2017.
- [7] C. P. Steinmetz, "On the law of hysteresis", *Proc. IEEE*, vol. 9, no. 2, pp. 3-64, 1892.
- [8] J. Reinert, A. Brockmeyer, and R. W. De Doncker, "Calculations of losses in ferro- and ferromagnetic materials based on modified Steinmetz equation," *IEEE Trans. Ind. Appl.*, vol. 37, no. 4, pp. 1055–1061, Jul./Aug. 2001.
- [9] J. Li, T. Abdallah, and C. R. Sullivan, "Improved calculation of core loss with nonsinusoidal waveforms," in *Proc. IEEE Annu. Meeting Ind. Appl. Soc.*, 2001, pp. 2203–2210.
- [10] K. Venkatachalam, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate prediction of ferrite core loss with non-sinusoidal waveforms using only Steinmetz parameters," in *Proc. IEEE Workshop Comput. Power Electron.*, 2002, pp. 36–41.
- [11] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 964–973, Feb. 2012.
- [12] H. Li, D. Serrano, S. Wang, T. Guillod, M. Luo, and M. Chen, "Predicting the B-H loops of power magnetics with transformer-based encoder-projector-decoder neural network architecture," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2023, pp. 1543–1550.
- [13] M. Chen, H. Li, D. Serrano, and S. Wang, "MagNet-AI: Neural network as datasheet for magnetics modeling and material recommendation," techRxiv, Preprint, 2023. [Online]. Available: <https://doi.org/10.36227/techrxiv.22726115.v1>.
- [14] T. Guillod, P. Papamanolis, and J. W. Kolar, "Artificial neural network (ANN) based fast and accurate inductor modeling and design," *IEEE Open J. Power Electron.*, vol. 1, pp. 284–299, 2020.
- [15] MagNet 2023 Team. Princeton University. (2023, Apr. 7). *2023 PEELS-Google-Tesla-Princeton MagNet Challenge*. [Online]. Available: <http://www.princeton.edu/~minjie/magnet.html>.
- [16] E. Colin Cherry, "The duality between interlinked electric and magnetic circuits and the formation of transformer equivalent circuits," *Proc. Phys. Soc. B*, vol. 62, no. 2, pp. 101–111, Feb. 1949.
- [17] Ion Boldea, "Linear electric machines drives, and MAGLEVs handbook", CRC Press, Taylor & Francis, New York, 2013.
- [18] K. A. Macfadyen, "Vector permeability," *Proc. Inst. Elect. Engineers*, vol. 95, issue 85, pp. 63-64, Jan. 1948.
- [19] Laithwaite E R, "Magnetic equivalent circuits for electrical machines", *Proc. Inst. Elect. Engineers*, vol. 114, no. 11, pp. 1805–1809, Nov. 1967.
- [20] M. Cheng, W. Qin, X. Zhu and Z. Wang, "Magnetic-Inductance: Concept, Definition, and Applications," *IEEE Trans. on Power Electron.*, vol. 37, no. 10, pp. 12406-12414, Oct. 2022.
- [21] W. Qin, M. Cheng, Z. Wang, Z. Ma, S. Zhu, M. Gu, X. Zhu, W. Hua, "Vector magnetic circuit theory and its preliminary applications," *Proc. CSEE*, 2023, <https://link.cnki.net/urlid/11.2107.tm.20231103.1033.002>.
- [22] Stoll R L. *The analysis of eddy currents*. Oxford: Oxford University Press, 1974: 1-69.
- [23] D. Serrano, H. Li, S. Wang, T. Guillod, M. Luo, etc., "Why MagNet: quantifying the complexity of modeling power magnetic material characteristics," *IEEE Trans. on Power Electron.*, vol. 38, no. 11, pp. 14292-14316, Nov. 2023.
- [24] P. M. Gradzki, M. M. Jovanovic, and F. C. Lee, "Computer aided design for high-frequency power transformers," in *Proc. IEEE APEC'90*, 1990, pp. 336–343.
- [25] R. Severns, "HF-core losses for nonsinusoidal waveforms," in *Proc. HFPC*, 1991, pp. 140–148.
- [26] M. K. Kazimierczuk, *High-Frequency Magnetic Components*, 2nd ed., New York, USA: John Wiley & Sons. Ltd., 2014, pp. 383–411.

Physics-Inspired Multimodal Feature Fusion Cascaded Networks for Data-Driven Magnetic Core Loss Modeling

Wei Xu, Jiyao Wang, Youkang Hu, Jing Xu, Zhongqi Shi

Abstract- This paper proposes a physics-inspired multimodal feature fusion cascaded network (PI-MFF-CN) for data-driven magnet core loss modeling based on MagNet database. The proposed methodology consists of a physically inspired network model and a multimodal feature fusion network model. The physically-inspired network is based on micromagnetism and the associated Landau-Lifshitz-Gilbert (LLG) equation, and is trained by embedding physical constants in the gradient learning process of the network. Meanwhile, a multimodal feature fusion method proposed in this paper combines the advantages of convolutional neural network (CNN) and FCNN to learn mixed sequence scale data. In this paper, the proposed model is trained and optimized based on the MagNet database, and validation experiments with a large number of materials are conducted. The results show that the proposed method can accurately predict the core loss with good generalization and robustness.

I. INTRODUCTION

Magnetic components such as inductors and transformers play a key role in power electronic systems. The design of these components is critical to improving system efficiency, lightweight construction and reducing energy consumption. Core loss of magnetic components is one of the key design parameters. However, current related research lags behind power devices and circuit topologies. Core loss modeling of magnetic materials involves multiple challenges such as different excitation waveforms, temperature, frequency, DC bias and different materials. Thus, the related studies of the rapid and precise core loss modeling methods in power magnetics are beneficial and imperative to the entire power electronics field.

Currently, the main standard methods of modeling core losses rely on curve-fitting, involving Steinmetz Equation (SE), modified Steinmetz Equation (MSE), improved generalized Steinmetz Equation (iGSE), and other SE-derived methods. These quasi-analytical-empirical methods provide reasonable estimates within a certain range and are most extended among manufacturers, but have obvious limitations when facing practical complex applications, such as ignoring cross-coupling effects, requiring material parameters, and accuracy limitations under specific waveforms.

Recent years have seen great advances in neural networks (NNs). NNs have proved effective in solving strong nonlinear mapping problems and exhibit a fast prediction speed. NNs have been widely applied for modeling core loss and integrated into optimization designs of magnetic components. From the perspective of modal information types used in NNs, there are two main technical ways of modeling the core loss: scalar-to-

scalar and sequence-to-scalar. In [6], the fully connected neural network (FCNN) with four scalar inputs is proposed to predict the core loss under specific types of $B(t)$ waveform shapes and operating conditions. Transfer learning for FCNN is employed in [7] to reduce the amount of training data required in core loss prediction when converting between different waveform shapes. These methods are usually limited to single-modal information input, and it is difficult to fully capture the core loss of the material relying on only scalar or sequence information. Sequence and scalar information are equally important in data-driven modeling. Sequence information typically contains temporal features and spatiotemporal relationships, while scalar data may contain critical global feature. Therefore, mixing multimodal information is crucial for more comprehensive core loss prediction and good generalization.

In addition, considering the voltage is usually imposed in power converters, $B(t)$ is more relevant and accessible than $H(t)$. Thus, for ease of practical application, both in the traditional equation-based and NNs-based core loss models, $H(t)$ is restricted to not be used as input. However, an additional $H(t)$ waveform can further enrich the hybrid multimodal features, which help improve the generalization and robustness of core loss modeling. Existing NN implementations require large amounts of data. Thus, it is a more economical method to use a network model with physically interpretable and small sample dependence to provide sequence information that contain features similar to real $H(t)$ to assist core loss modeling.

Based on the above researches and discussion, this paper presents a physics-inspired multimodal feature fusion cascaded network (PI-MFF-CN) for data-driven magnet core loss modeling. The rest of this article is organized as follows. Section II describes the framework of the proposed modeling method and explains how to construct the prediction mechanism. In Section III, the training and optimization procedure of the proposed data-driven magnetic core loss modeling based on MagNet database is described in detail. Finally, Section IV concludes this paper.

II. METHODOLOGY

This section presents the framework of the proposed modeling method, namely PI-MFF-CN in detail and explains how to construct the prediction mechanism. Firstly, this section introduces the basic theory of micromagnetism and LLG Equation, as well as discusses and analyzes the basic network structures FCNN and CNN.

A. Basic Theory of Micromagnetism and LLG Equation

The theory of micromagnetism serves as a crucial link between macroscopic and microscopic analyses of magnetic materials. In micromagnetism, the complex magnetic properties of multiple atoms are effectively simplified into a single magnetic moment. By delving into the interactions among these magnetic moments, micromagnetism offers a microscopic viewpoint, allowing for the observation of the dynamic response across the entire magnetic material. Micromagnetic simulations prove invaluable for comprehending the magnetic characteristics of materials, fine-tuning material parameters, and predicting material performance under diverse conditions.

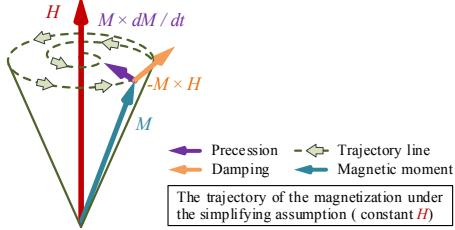


Fig. 1. The magnetization trajectory described by LLG equation.

The primary theoretical equation in micromagnetism is the Landau-Lifshitz-Gilbert (LLG) equation, which describes the motion of the magnetic moment, as expressed by:

$$\dot{m}(r,t) = -\gamma m(r,t) \times H_{\text{eff}} + \alpha m(r,t) \times \dot{m}(r,t) \quad (1)$$

where m is the unit magnetization vector and \dot{m} is the first-order derivative to time. γ , α are the gyromagnetic ratio and the Gilbert damping constant respectively. Fig. 1 shows the magnetization trajectory described by LLG equation. The effective field H_{eff} contains the energy contributed by the following common interactions, as shown in:

$$H_{\text{eff}} = H_{\text{ext}} + H_{\text{exch}} + H_{\text{ani}} + H_d \quad (2)$$

where H_{ext} is the external field, H_{exch} is the exchange field, H_{ani} is the anisotropy field, and H_d is the demagnetization field. These interacting fields compete with each other in the magnetic system to keep the magnetic moment moving and eventually reach an equilibrium state. In these effective fields, there are coefficients associated with the intrinsic characteristics of the material, such as the coupling constant A , magnetic anisotropy constant K , and saturation magnetization M_s . These coefficients play a pivotal role in shaping the magnetic response of the material. The coupling constant A influences the degree of interaction between neighboring magnetic moments, the magnetic anisotropy constant K characterizes the material's orientation preferences in different directions, and the saturation magnetization M_s dictates the material's maximum achievable magnetization.

Through the comprehensive analysis of the overall magnetic moment variation within magnetic material, the $B(t)$ waveform of different materials can be simulated from $H(t)$ waveform. Considering the limitation that $H(t)$ waveforms cannot participate in the actual core loss modeling process, it is important to build the inverse model ($B(t) \rightarrow H_{\text{LLG}}(t)$) of the LLG Equation.

B. Basic Theory of FCNN and CNN

FCNN consists of an input layer, multiple hidden layers, and an output layer. Fig. 2 depicts a simple example of FCNN. Neurons in each layer are connected to all neurons in both the

previous and the following layers. This mode of fully connected association enables information to flow freely through the network, allowing for the learning of the hidden, complex, nonlinear relationship. In the forward propagation process, each neuron consists of a linear transformation h and a nonlinear transformation f . In training, the loss function evaluates output-target error, and backpropagation with gradient descent updates weights and biases to minimize the loss.

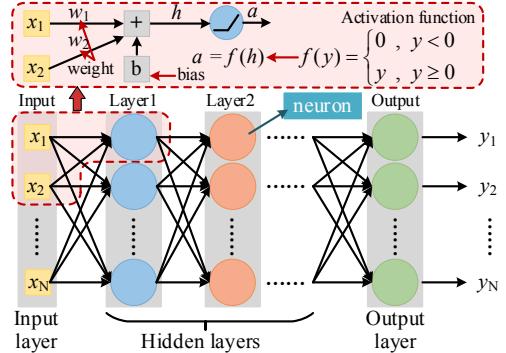


Fig. 2. A simple example of FCNN.

While FCNN exhibits strong global learning, the growing number of layers leads to an exponential rise in parameters, potentially causing slow learning and overfitting. Additionally, the fact that FCNN treats each input equally also makes it unsuitable for direct learning data with hybrid multimodal sequence-scalar information.

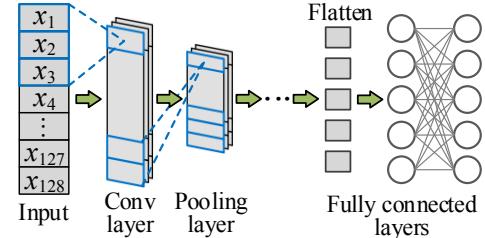


Fig. 3. A simple example of CNN.

Different from FCNN, a typical CNN is constructed by convolutional layers, pooling layers, and fully connected layers, as shown in Fig. 3. A convolutional layer extracts feature maps by sliding filter kernels and computing convolutions between input local regions and kernels. Since the convolution operation itself is linear, the ReLu activation function is used to introduce a nonlinear mapping to help the network learn more complex features.

Pooling is a form of subsampling after convolution that is used to reduce the feature dimension. After several alternate convolutional and pooling layers, one or more fully connected layers are usually used to map the features learned by the previous layers. CNN reduces the dimension of sequence information by extracting local key features with translation invariance. However, processing mixed sequence and scalar information may lead to loss of key scalar information.

Thus, this article combines the advantages of CNN and FCNN to learn hybrid sequence-scale data through special network structure designs.

C. Framework of Magnet Core Loss Modeling

As shown in Fig.3, the proposed methodology consists of two cascaded sub-models: the physics-inspired network model (sub-model A) and the multimodal feature fusion network model (sub-model B). The aim of sub-model A is to provide new sequence information ($H_{LLG}(t)$) for the next cascaded sub-model B to solve the limitation that $H(t)$ waveforms cannot

participate in the actual prediction process. In addition, through embedding physical constants (A, K, M_s) into the learning process of the specially designed networks, the sub-model A can be regarded as the inverse model ($B(t) \rightarrow H_{LLG}(t)$) of the LLG Equation with physical interpretability.

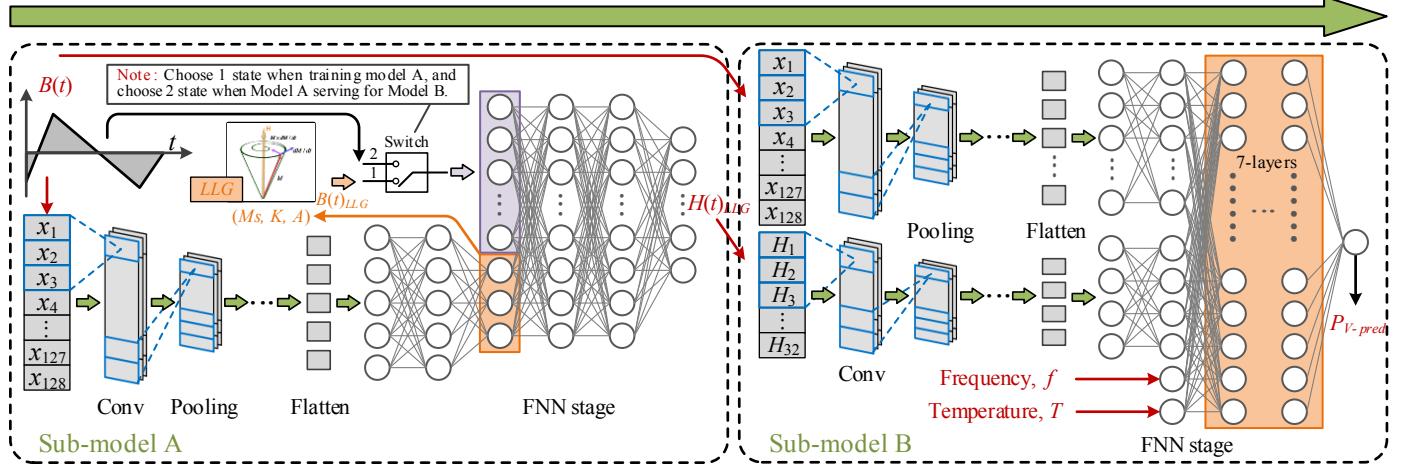


Fig. 4. The framework the proposed PI-MFF-CN method.

The flowchart of the proposed physics-inspired network (sub-model A) is shown in the left half of Fig. 4, including one CNN stage and one FCNN stage. This paper utilizes the open-source micromagnetic framework, MagTense, for a micromagnetic simulation to inspire the network learning of the inverse model ($B(t) \rightarrow H_{LLG}(t)$) of the LLG Equation. MagTense applies the finite difference method to discretize the LLG equations into difference equations solved iteratively. The material space is divided into grids, modeling the evolution of magnetic moments at these points. By calculating the magnetic moments' evolution at each grid point, the magnetic flux density $B_{LLG}(t)$ is converted from the final overall magnetic moments M in the material. The key unknown constants (M_s, K, A) in micromagnetic simulation are obtained from CNN's outputs.

In the training process, $B(t)$ waveforms are input into CNN to extract features and infer material characteristic coefficients (M_s, K, A). Before entering the next FCNN stage, MagTense simulates $B_{LLG}(t)$ waveform based on the coefficients (M_s, K, A) from the previous iteration and the current $H(t)$ waveform. Then $B_{LLG}(t)$ and are appended to the outputs of CNN and fused into the global learning stage of FCNN to predict $H_{LLG}(t)$. The loss function is used to evaluate the error between the $H_{LLG}(t)$ and the target value $H(t)$. As the weights and biases of the whole networks (CNN & FCNN) are updated through the backpropagation and gradient descent to minimize the loss, the material characteristic coefficients (M_s, K, A) tend to shape the magnetic response of the material. Thus, $B_{LLG}(t)$ is gradually becoming similar to the actual $B(t)$ waveform.

After training, the inverse model at this state can be expressed as the regression from $(B(t), B_{LLG}(t))$ to $H_{LLG}(t)$, $H(t)$ is still needed in the application to calculate the $B_{LLG}(t)$. Considering that $B_{LLG}(t)$ is already similar to $B(t)$, and $H_{LLG}(t)$ is designed as an intermediate information in the proposed cascaded network. Thus, $B_{LLG}(t)$ is replaced by $B(t)$ when sub-

model A serving for sub-model B. Although this change may cause the slight errors in $H_{LLG}(t)$ waveform, the hidden deep features remain unchanged and can be mined through later CNN stage, which is helpful to the core loss modeling. The inverse model at this state can be expressed as the regression from $B(t)$ to $H_{LLG}(t)$.

Benefit from sub-model A, more diverse hybrid multimodal sequence-scalar information is help to improve the generalization and robustness of core loss modeling. Sequence data typically contains temporal information and spatiotemporal relationships, while scalar data may contain critical global information. Since the complex relationship between sequence and scalar data, effective feature fusion is challenging and important to improve the accuracy of core loss modeling. The flowchart of the proposed multimodal feature fusion network (sub-model B) is shown in the right half of Fig. 4, including two parallel CNN stage branches and one FCNN stage. The output $H_{LLG}(t)$ of the sub-model A is cascaded to the input of the upper CNN stage in sub-model B to form a cascaded network.

After extracting the deep feature mappings from the sequence information ($B(t), H_{LLG}(t)$) through the upper and lower CNN branches, the obtained feature mappings are appended to the remaining scalar information (temperature: T , frequency: f), and then fused into the global learning stage of FCNN for more accurate core loss prediction.

The advantage of this strategy is utilizing the advantages of CNN for sequence feature mapping and then learning together with other scalar data by FCNN. Through this multimodal feature fusion method, the sub-model B can more comprehensively learn the deep associations in mixed sequence-scalar data to accurately predict magnetic core losses with good generalization.

III. TRAINING AND OPTIMIZATION OF DATA-DRIVEN MODELS

To validate the effectiveness of magnetic core loss modeling based on the proposed PI-MFF-CN method, the related network models shown in Fig. 4 are built using Python language with the PyTorch library. Next, this section describes the detailed training and optimization process of data-driven models.

A. MagNet Database

Training an accurate PI-MFF-CN model requires first and foremost an extensive and reliable amount of data, more specifically, $B(t)$ and $H(t)$ waveforms under different conditions. This paper uses a large-scale open-source power magnetic materials database - MagNet (mag-net.princeton.edu), provided by the MagNet Challenge 2023 Project. MagNet database measures the core loss by the four-wire voltage/current method and in its current state contains a variety of ferrite materials in the 50 kHz to 500 kHz, 0 °C to 120 °C, and 10 mT to 300 mT range for sinusoidal, triangular, and trapezoidal waveforms.

Further details on the data acquisition system of MagNet can be found in [16]. In this work, the material N87 is used to validate the proposed data-driven core loss modeling method, and the rest materials will be used for further analysis of the model generalization and robustness. Before the training process, the dataset of material N87 was randomly divided into training and test sets at an 8:2 ratio.

B. Data Normalization

In the feature vectors, different physical variants possess different units such as temperature T and frequency f. Scale differences of different modal features may cause certain local features to have a greater impact on weight updates during model training, which is detrimental to overall stability and generalization. In this paper, the linear scaling method is used to transform the input features to a suitable range.

C. Training of Sub-Model A

To obtain the inverse model of the LLG equation, the physics-inspired network model (sub-model A) is first for training. In this section, mean squared error (MSE) is applied for the error evaluation of predictions, the MSE is defined by:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (6)$$

where y_i is the i -the tagged value, \bar{y}_i is the i -th predicted value, and n is the number of samples. After a few epochs' iterations, the coefficient (A, K, Ms) and MSE gradually becomes stable. This sub-model A can well meet the design expectations for further enrich the hybrid multimodal features, and help to improve the generalization and robustness of core loss modeling.

D. Cascading and Training Settings of Sub-Model B

After obtaining the physics-inspired inverse model (sub-model A) of LLG equation, the output HLLG(t) of the sub-model A is cascaded to the input of the upper CNN stage in sub-model B to form a cascaded network model for magnetic core loss modeling. When training the model, sequence data ($B(t)$) and Scalar data (T, f) are input to the cascaded network for

prediction. The loss function evaluates the error between the predicted value and the true value, and then updates the weights of the sub-model B through reverse gradient propagation to minimize the error. It should be noted that the gradient does not update sub-model A that the network parameters are fixed.

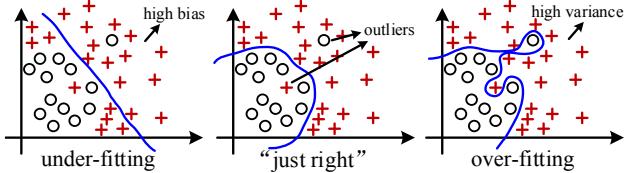


Fig. 5. Different fitting states of the model during training.

Before the training process, considering magnetic core loss measurements may be affected by noise or uncertainty, there will be outliers that deviate from the normal range. Since MSE gives more weight to outliers, it will sacrifice the prediction effect to other normal data points in the material N87 database, leading to a degradation of the overall model performance, i.e., the model performs “too well” on the training set (over-fitting) while “poorly” on the test set (under-fitting). Fig. 5. Shows different fitting states of the model during regression. Therefore, it is necessary to select a more suitable loss function in the gradient updating of the sub-model B for core loss modeling. The mean absolute error (MAE) is another commonly used loss function that is insensitive to outliers:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (7)$$

Different from the mean reversion principle of MSE, MAE based on median regression is more robust to outliers. However, it is not negligible near the zero point and will be difficult to converge when the gradient is too large. In order to combine the advantages of MSE and MAE to improve the robustness and generalization performance of the core loss model, this paper adopts the hybrid method $SmoothL_1$ loss function, which is defined by:

$$SmoothL_1 = \begin{cases} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (y_i - \bar{y}_i)^2 & , |y_i - \bar{y}_i| \leq 1 \\ \frac{1}{n} \sum_{i=1}^n (|y_i - \bar{y}_i| - \frac{1}{2}) & , |y_i - \bar{y}_i| > 1 \end{cases} \quad (8)$$

where δ is a hyperparameter. The $SmoothL_1$ loss function achieves two goals: (1) Weaken the over-sensitivity problem of outliers (2) The property of being globally derivable make the model converge faster than MAE .

In addition, as the model's complexity increases, so does its learning depth, making it still easily prone to over-fitting. In this paper, ridge regression regularization is added to the training process of sub-model B to decrease the weight of features containing useless information, thereby reducing the risk of overfitting. Ridge regression is done by adding L2 regularization after the $SmoothL_1$, as follows:

$$J = \min J(\theta) = SmoothL_1 + \lambda \cdot \sum_{j=1}^n \theta_j^2 \quad (9)$$

where $\lambda \cdot \sum \theta_j^2$ is the regularization term, θ is network parameters, and λ is the regularization coefficient. Note that setting λ too

large will make it approach zero, causing under-fitting. Thus, it is important to choose an appropriate regularization coefficient λ .

E. Adaptive Adjustment Learning Rate of Sub-Model B

For improving efficiency and saving training time, this paper chooses a hybrid optimization strategy of adaptive adjustment learning rate combined with random grid search for hyperparameters. The learning rate of the model determines the scale factor of the gradient. A too-high learning rate will cause the optimizer to exceed the optimal value, while a too-low learning rate will result in a longer training time. It's hard to find a static, effective, unchanging learning rate. ReduceLROnPlateau optimizer in Pytorch is chosen to monitor the SmoothL1 loss. If the loss does not improve within p_a patience epochs, the learning rate l_r is reduced by a predefined factor f_b to slow down l_r . Based on the above strategy, the final sub-model B with optimal hyperparameters is obtained.

F. The Numbers of Parameters of Model

Material	A	B	C	D	E
Number of Parameters	139938	139938	139938	139938	139938
Model size / KB	(176+385)	(176+385)	(176+385)	(176+385)	(176+385)

IV. CONCLUSION

This paper presents a physics-inspired multimodal feature fusion cascaded networks for data-driven magnet core loss modeling based on MagNet database. In this work, a physics-inspired network based on micromagnetism and related Landau-Lifshitz-Gilbert (LLG) equation is proposed to provide new sequence information ($H_{LLG}(t)$) for the next cascaded core loss prediction model to solve the limitation that $H(t)$ waveforms cannot participate in the actual prediction process. Through embedding physical constants (A , K , M_s) into the gradient learning process of the network, the physics-inspired network can be regarded as the inverse model ($B(t) \rightarrow H_{LLG}(t)$) of the LLG Equation with physical interpretability. Meanwhile, this model combines the advantages of CNN and FCNN to learn hybrid sequence-scale data, i.e., sequence feature mapping is performed by a pair of parallel CNN branches, and then the mapping is concatenated with other scalar data into FCNN for global learning. This method can accurately predict the core loss with good generalization and robustness.

REFERENCES

- [1] 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," *IEEE Trans. Power Electron.*, vol. 31, no. 11, pp. 7909-7925, Nov. 2016.
- [2] J. Wang, S. Zhang, C. Li, L. Wu, and Y. Wang, "A Data-driven method with mode decomposition mechanism for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Power Electron.*, vol. 37, no. 11, pp. 13684-13695, Nov. 2022.
- [3] D. Serrano et al., "Why MagNet: Quantifying the complexity of modeling power magnetic material characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 11, pp. 14292-14316, Nov. 2023.
- [4] H. Li et al., "How MagNet: Machine learning framework for modeling power magnetic material characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15829-15853, Dec. 2023.
- [5] H. Li, D. Serrano, S. Wang and M. Chen, "MagNet-AI: Neural network as datasheet for magnetics modeling and material recommendation," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15854-15869, Dec. 2023.
- [6] N. Rasekh, J. Wang, and X. Yuan, "Artificial neural network aided loss maps for inductors and transformers," *IEEE Open J. Power Electron.*, vol. 3, pp. 886-898, Nov. 2022.
- [7] H. Li et al., "MagNet: An opensource database for data-driven magnetic core loss modeling," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2022, pp. 588-595.
- [8] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A Data-driven Auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3478-3487, May 2021.
- [9] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439-453, 2018.
- [10] F. Wang, R. Liu, Q. Hu, and X. Chen, "Cascade convolutional neural network with progressive optimization for motor fault diagnosis under nonstationary conditions," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2511-2522, Apr. 2021.
- [11] E. Peker and A. Wiesel, "Fitting generalized multivariate SmoothL1 loss functions," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1647-1651, Nov. 2016.
- [12] G. Yating, W. Wu, L. Qiongbin, C. Fenghuang and C. Qinjin, "Fault diagnosis for power converters based on optimized temporal convolutional network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-10, Sep. 2021.
- [13] N. Buslim, I. L. Rahmatullah, B. A. Setyawan and A. Alamsyah, "Comparing bitcoin's prediction model using GRU, RNN, and LSTM by hyperparameter optimization grid search and random search," in *Int. Conf. Cyber IT Serv. Manag.*, 2021, pp. 1-6.
- [14] V. G. Bar'yakhtar and B. A. Ivanov, "The Landau-Lifshitz equation: 80 years of history, advances, and prospects," *Low Temp. Phys.*, vol. 41, no. 9, pp. 663-669, Sep. 2015. LLG
- [15] Z. Zhu, Y. Sun, Q. Zhang, and J.-M. Liu, "Dynamics and scaling of low-frequency hysteresis loops in nanomagnets," *Phys. Rev. B*, vol. 76, no. 1, p. 014439, Jul. 2007. LLG
- [16] R. Bjørk, E. B. Poulsen, K. K. Nielsen and A. R. Insinga, "MagTense: A micromagnetic framework using the analytical demagnetization tensor," *J. Magn. Magn. Mater.*, vol. 535, pp. 168057, Oct. 2021. Magtense.

Magnetization Mechanism-Inspired Neural Networks for Core Loss Estimation

Qiujie Huang¹, Yang Li², Yu Dou³, Bo Li³, and Sinan Li¹

¹School of Electrical and Computer Engineering, University of Sydney, Australia

²Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

³Department of Electrical Engineering, Hebei University of Technology, Tianjin, China

Abstract- Fast, accurate, and cost-effective core loss modelling is critical to achieve the design of high-density high-efficiency power electronics systems. However, conventional core loss modelling methods often demonstrate a significant lack of accuracy, design efficiency and/or cost-effectiveness especially under complex operating conditions. To tackle this problem, this document proposes a new core loss modelling method, Magnetization Mechanism-Inspired Neural Network (MMINN). MMINN is a hybrid data-driven and physics-driven model. It is designed to capture the fundamental magnetization process of magnetic materials at the micro-level. Therefore, MMINN can not only achieve high accuracy and computational efficiency as many other fully data-driven models do, but also enjoy the simplicity (fewer parameters to tune) and generality of a purely physic-driven model. Model training and testing are based on the hysteresis loops of five ferrite materials in the open-source core loss database - MagNet. Experimental results prove the accuracy and the reliability of the proposed model, making it highly applicable for power magnetics design. Additionally, a user interface is provided to visualize model prediction data.

I. INTRODUCTION

Power magnetics often have the largest volume and primary source of power losses, posing significant challenges to high-density and high-efficiency power electronics design [1]-[2]. Modelling of power losses, especially core losses, is a fundamental task of high-performance power magnetics design. The trend of miniaturization of power electronic devices makes fast, accurate, and low-cost core loss modelling imperative. However, modeling core losses is untrivial due to the highly nonlinear nature of the magnetization process of magnetic materials. Steinmetz empirical models are some of the most widely used models in the field of power electronics, which map the relationship between excitation and core losses through a few curve-fitting parameters [3]-[6]. Steinmetz models are simple to use but the key challenge is its lack of accuracy and generality for different core materials and/or operations under different conditions (e.g., variations in frequency, excitation profiles, temperature) [7]-[8]. Thus, it is time to upgrade the Steinmetz equation and advance the entire power electronics society's understanding of magnetization mechanisms.

To this end, extensive research has been devoted to improving core loss modelling by reconstructing the dynamic hysteresis process (i.e., the hysteresis loops of magnetic field strength H and flux density B), which provides deeper physical insights at the material level. Different dynamic hysteresis models can be divided into three main categories:

- **Fully physics-driven models** – The models are mainly based on the microscopic (magnetic domain) energy calculation and traditional numerical approaches, where the Preisach model [9] and the J-A model [10] are the most well-documented branches in literature. Such models are physically interpretable and therefore very accurate [9]-[14]. Nevertheless, they still suffer from the following problems: (i) high computational complexity due to iterative calculations and (ii) complex parameter identification process.
- **Fully data-driven models** – The models are based on Machine Learning (ML) techniques, especially general-purpose neural networks (NNs) [15]-[19]. Disadvantages of fully data-driven model includes (i) a large dataset which are generally costly and time-consuming to obtain and (ii) a large model size to ensure a satisfactory level of prediction accuracy, thus expensive computational unit for model training. Also, the black box nature makes it (iii) difficult to understand the decisions made inside the model.
- **Hybrid models** – Hybrid models combine data-driven and physics-driven characteristics to alleviate their inherent shortcomings. The hybrid model is the most promising method for core loss modelling, owing to (i) significantly reduce the training data requirements, (ii) show a simpler structure, (iii) have lower memory requirements and (iv) provide some physical insights [20]-[22].

In this paper, a hybrid model, Magnetization Mechanism-Inspired Neural Network (MMINN), is proposed for power magnetics design. The proposed model uses separate subnetworks to reconstruct dynamic and static hysteresis components where the static subnetwork adopts the Prandtl-Ishlinskii (PI) stop operator to characterize the material's static magnetic moment changes, and the dynamic subnetwork adopts the idea of ladder network segmentation to account for the dynamic effects within the magnetic material. Based on measurement data from the MagNet database [17], the trained MMINN allows fast and accurate core loss estimation and magnetization process simulation under different operating conditions (temperatures, excitation frequencies, and excitation waveforms). Besides, a user interface (UI) platform is designed to seamlessly integrate with MMINNs, empowering users with a diverse range of interactive possibilities and providing insightful data visualization.

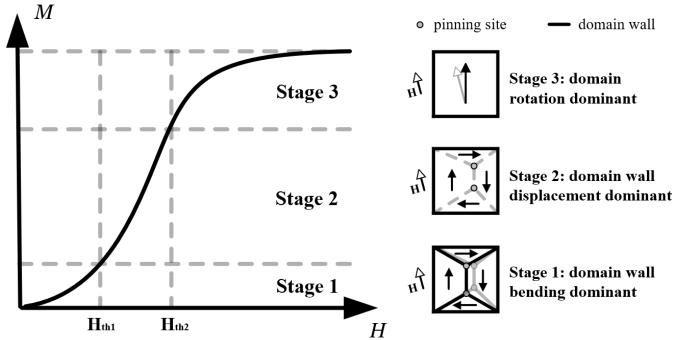


Fig. 1. Magnetization process (a) M-H curve of soft magnetic materials (b) Cubic soft magnetic material magnetization schematic.

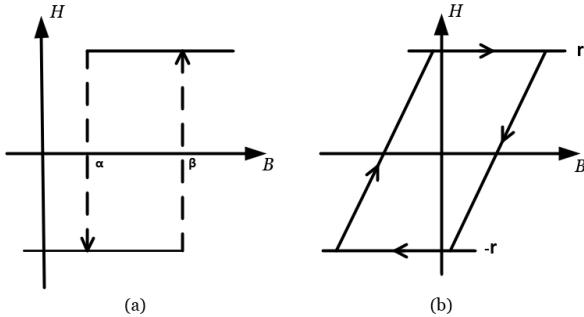


Fig. 2. (a) Preisach model: Relay hysteresis operator (b) PI model: Stop hysteresis operator.

This paper is organized as follows. Section 2 reviews underlying magnetization mechanisms and some physical models to simulate magnetization process. Section 3 introduces MMNN concept. Section 4 presents the research results for the MagNet Challenge 2023. Section 5 concludes the paper.

II. MAGNETIZATION MECHANISMS

The dynamic magnetization process of soft magnetic materials can be constructed through a B-H loop, and the core loss W can be expressed by the enclosed area $\int H dB$ as:

$$W = V(f \int H dB), \quad (1)$$

where V denotes the magnetic component's effective volume and f denotes the operating frequency. According to the loss separation theory [23]-[24], the core loss can be decomposed into the sum of a static hysteresis contribution W_h and a dynamic contribution W_{dyn} . Subsequently, in the dynamic hysteresis modeling, the magnetic response of the material to changes in external magnetic flux density is also composed of two parts. The overall magnetic field strength H can be equivalently interpreted as:

$$H = H_h + H_{dyn}. \quad (2)$$

A. Static component modeling – PI model

The hysteresis contribution H_h is considered a static term because it is independent of excitation frequency. Fig.1(a) illustrates the static magnetization process of a magnetic domain through the magnetization characteristic curve. Magnetic domain, the smallest unit showing magnetic

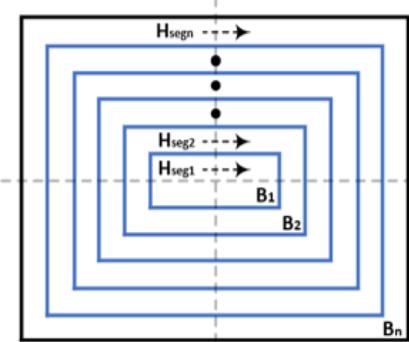


Fig. 3. Core cross-section with hypothetical eddy current paths.

characteristics, undergoes magnetic domain wall movement (stage 1: bending and stage 2: displacement) and magnetic domain rotation (stage 3) during the magnetization process as shown in Fig.1(b) [25]. The static magnetization characteristics of magnetic domain in different materials are dominated by these two magnetization mechanisms. And the B-H loop of a magnetic material is an aggregation of the magnetization status of each individual wall domain inside the material.

The hysteresis operator is often used to mimic/characterize the static magnetization process of a magnetic domain shown in Fig.1 (a). [26]. The relay hysteresis operator is a typical example having a rectangular hysteresis loop with two thresholds (α , β), as depicted in Fig.2(a). It can be seen as a rough description of domain rotation (stage 3) but does not account for the intermediate transition process (stage 1 and stage 2) [27]. To include all magnetization mechanisms and improve model accuracy, Prantl [28] introduced the stop hysteresis operator S_r , guaranteeing the continuity of B-H mapping (see Fig.2(b)). Physically, it provides an acceptable description of reversible and irreversible components of hysteresis. Its corresponding static B-H hysteresis model is called the PI model, which can be interpreted by:

$$H_h(B) = \sum_{r=1}^n \mu_r(f, T) S_r(B). \quad (3)$$

where n is the number of operators. Each stop operator S_r is characterized by a saturation level r , and the corresponding weight function μ_r represents its distribution proportion.

The challenge of PI models is computing the distribution function. Conventional methods include formulation of the Everett function and particle swarm optimization [29]-[30]. However, these methods are time-consuming and cannot guarantee model's reliability. Therefore, the proposed model introduces a feed-forward neural network (FNN) to represent μ_r of the PI model and ultimately describes the static component of the magnetization process (see Section III).

B. Dynamic component modeling – Core loss ladder network

The static hysteresis models will produce large core loss prediction errors at high frequencies [26], [31]. At this point, the dynamic hysteresis contribution H_{dyn} should be considered, which mainly originates from electromagnetic induction. Maxwell's equations can be theoretically applied to compute

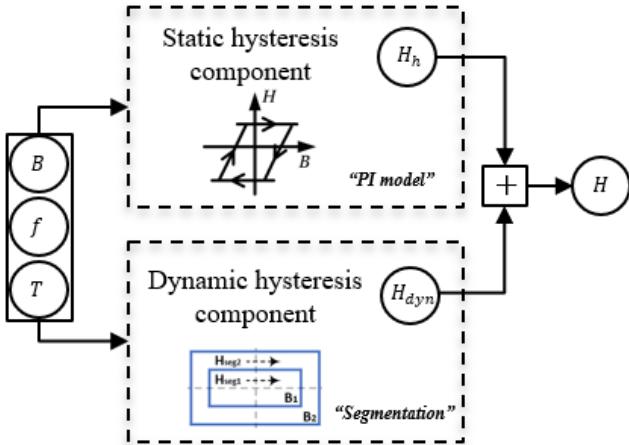


Fig. 4. Physical core loss model schematic diagram.

the overall magnetic field strength. However, in practice, there are differences between the calculated and measured magnetic fields. The differences are caused by anomalous loss behaviors, including localized eddy currents near the magnetic domain wall, spin damping, internal friction, and dimensional resonance effects [32]. These phenomena will lead to spontaneous magnetization vector rotation and circulating currents around the magnetic domain walls. Thus, Bertotti [23] separated H_{dyn} into classical eddy current losses and anomalous losses, as follows:

$$H_{dyn} = C_{cl} \frac{dB}{dt} + C_a \left| \frac{dB}{dt} \right|^{0.5}. \quad (4)$$

where C_{cl} is the classical eddy current coefficient and C_a is the anomalous loss coefficient. Both loss coefficients depend on the material and the component dimensions. Under low-medium frequency excitation, both loss coefficient can be treated as constant and identified by the core loss separation procedure [34]. However, under high frequency excitation, they become multiple factors dependent, including temperature effect, dimensional resonance, and skin effect [35]. Consequently, dynamic magnetization components computation at high frequencies become a popular research topic.

Zhu et al. [14] proposed a ladder network for magnetic cores to generalize the hysteresis model to high-frequency applications. This model regards the skin effect of magnetic field inside the magnetic core as the main reason for the non-uniform flux density in the high frequency environment. To account for the skin effect, the cross-section of the magnetic core is subdivided into several theoretical paths for eddy currents in this model (see Fig.3). The magnetic flux in each segment is regarded as uniformly distributed, so that Eq. (4) can be applied to describe the dynamic hysteresis effects of each segment. The idea of segmentation enables the dynamic magnetization process to be parameterized over a wide frequency range. Hence, the proposed model will apply an RNN to mimic section segmentation, where each hidden neuron reflects the dynamic effect of each hypothesized eddy path (see Section III).

III. MAGNETIZATION MECHANISM-INSPIRED NEURAL NETWORK

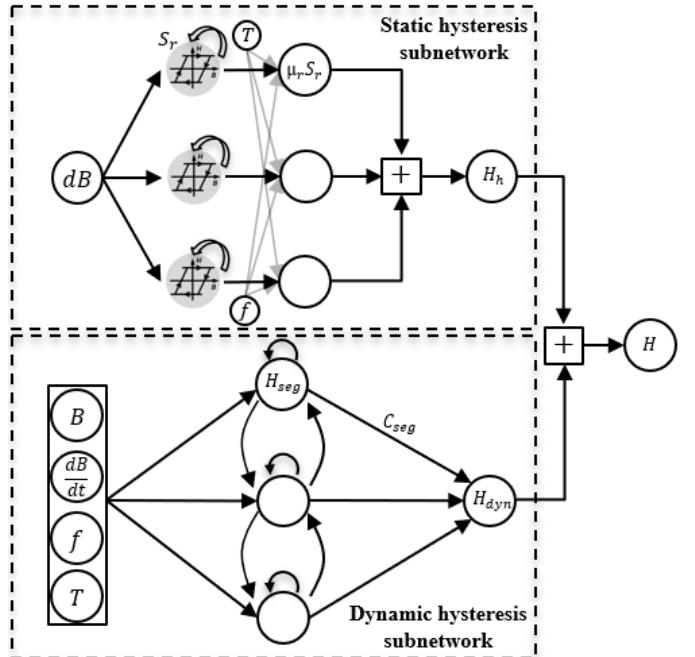


Fig. 5. Magnetization mechanism-inspired neural network.

This paper aims to characterize core loss information of soft magnetic material. The proposed model is inspired by magnetization mechanism and parameterizes the equivalent field strength in the form of $H(B)$ to characterize the dynamic magnetization process. Finally, the core loss prediction is obtained according to Eq. (1).

A. Model architecture

The magnetization process involves the rearrangement of the magnetic moments of the internal magnetic domains, consisting of static and dynamic hysteresis components as in Eq. (2). Based on the physical model introduced in section II, a physical core loss model can be constructed as shown in Fig. 4.

Analogously, to build a hybrid model, two sub-neural networks are used to embed the physical models and characterize static and dynamic hysteresis component. Fig.5 depicts the model structure of magnetization mechanism-inspired neural networks.

Static magnetization occurs when the magnetic moments have enough time to respond to an applied flux density. The change in magnetic moment caused by domain rotation and domain wall motion contains reversible and irreversible components. As discussed in Section 2, the PI stop operator can describe reversible and irreversible energy changes, so the static hysteresis subnetwork is constructed based on the PI model and Eq. (3). As shown in Fig. 5, the stop operators' response S_r will be treated as the inputs of the FNN in the static hysteresis subnetwork. Additionally, operating frequency and temperature are auxiliary inputs. Next, the FNN outputs $\mu_r S_r$, that is, the magnetic field strength generated by each type of stop operator (with the same saturation level). From the physical perspective of magnetization, the magnetization of materials is based on the magnetic moment changes of all magnetic domain, where each magnetic domain is modeled by a stop operator. Therefore, we

can estimate H_h by summing induced magnetic field strengths of all operators.

The subnetwork used to model the dynamic hysteresis component will adopt the idea of section segmentation in the core loss ladder network, as mentioned in Section 2.3, to transform the complex nonlinear behavior into a combination of linear problems. From Eq. (4), we can rearrange the overall field strength due to the dynamic effects as:

$$H_{dyn} = \sum_{k=1}^m C_{seg_k} * g\left(\frac{dB_{seg_k}}{dt}\right), \quad (5)$$

where m is the number of the segments, C_{seg} is the dynamic effects coefficient of a specific segment, containing the information of the material property information. $g(\cdot)$ is a nonlinear function with sequential dependency characteristic. We model the dynamic contribution using an RNN, where the number of hidden layer neurons in the RNN represents the number of segments. The neuron hidden state represents $g\left(\frac{dB_{seg}}{dt}\right)$, which is depicted as H_{seg} in Fig.5. The subnetwork learns C_{seg} from the training data and express it as the weight from the hidden layer to the output. Eventually, the dynamic hysteresis component is the weighted sum of the magnetic field strengths generated in all hypothetical segments.

Mathematically, the operation of MMINN can be written as:

$$\begin{cases} x(t) = {}^1w^1v(t) \\ h(t) = {}^2w^2v(t) + {}^hwH_{seg}(t-1) \\ H_{seg}(t) = f_h[h(t)] \\ H_{dyn}(t) = C_{seg}H_{seg}(t) + b \\ H_h(t) = \sum f_h[x(t)] \\ H(t) = H_h(t) + H_{dyn}(t) \end{cases}, \quad (6)$$

where 1w and 2w are the weight of input nodes of the two subnetworks, ${}^1v(t)$ and ${}^2v(t)$ are the inputs of networks, $x(t)$ is the output of the input layer of the static hysteresis subnetwork, hw is the weight of the hidden nodes, b is the bias of the output node of the dynamic hysteresis subnetwork and $f_h[\cdot]$ is the activation function of the hidden layer. Here we choose sigmoid function as the activation function.

B. Model training

Before model training, all data needs to be preprocessed, which ensures that network parameters are learned on a similar scale and reduces training complexity. This process includes variable transformation into log scale (frequency) and standardization (all input variables).

For the main model training process, the objective function plays a crucial role in the neural network's performance and defines the direction of model parameters optimization. Here, we use mean square error metric to constrain the predicted field strength H to approximate to the true field strength \hat{H} , as follows:

$$L = \frac{1}{n} \sum_{i=1}^n (H_i - \hat{H}_i)^2, \quad (7)$$

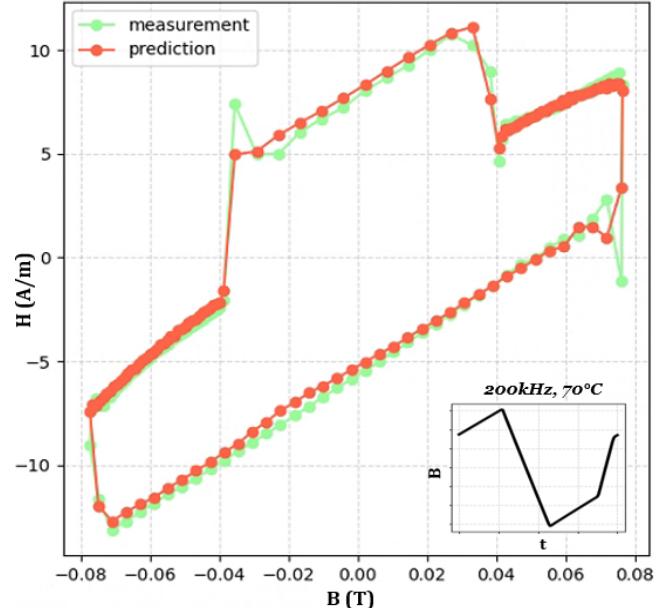


Fig. 6. Magnetization process prediction of material B under a trapezoidal excitation.

Table. 1. Number of parameters of each model.

Material	A	B	C	D	E
Num. of θ	1084	1084	1084	1084	1084
File size	7kb	7kb	7kb	7kb	7kb

where n denotes the number of the training data. Then, the proposed framework applies Adam [36] as the main optimization algorithm, which combines the benefits of adaptive gradient descent and momentum methods. The parameters related to the Adam algorithm will be improved through the performance of the model on the validation dataset.

IV. RESULTS AND DISCUSSION

The model results and discussions in this section are based on the problem setting of the MagNet challenge 2023. There are five unknown ferrite materials datasets for core loss model development, with each material containing between 580 and 7400 data points including different excitation waveforms, temperatures, and excitation frequencies.

A. Model size

The number of operators and the number of core cross-section segments (i.e., the number of RNN neurons in the dynamic hysteresis subnetwork) are the main factors that determine the performance and generalization ability of the model. They are defined as hyperparameters of the model. For the hyperparameters selection of the following network instances developed, we adopt the Optuna hyperparameter optimization framework, which is an improved version of the grid search technique [37].

MMINN is a hybrid model that naturally has advantages in scale over fully data-driven models. Therefore, instead of focusing on the compactness of the model, the optimization of

Table. 2. MMINN core loss prediction error based on 5 material training datasets.

Material	A	B	C	D	E
Abs. Avg. of Relative Error	3.26%	1.40%	1.73%	1.39%	1.80%
Abs. 95% of Relative Error	8.68%	3.67%	4.68%	3.58%	4.89%
Abs. 99% of Relative Error	14.3%	5.43%	8.29%	4.79%	7.17%
Abs. Max of Relative Error	26.6%	19.0%	24.3%	8.17%	12.9%

hyperparameters should be more focused on making core loss prediction reach the (average) relative error accuracy (<5%) acceptable for engineering applications. We select material A with a relatively suitable amount of data for hyperparameter optimization. Within 1000 training epochs, we found that 30 operators and 30 segments can achieve an average relative error of 5% for core loss prediction. Small adjustments (adding 1-2 neurons) have little impact on model accuracy, but huge adjustments (adding 10 neurons) will double the model size. Thus, we set this set of hyperparameters as the baseline for all material models. Table 1 lists the number of model parameters.

B. Model accuracy

The model accuracy shown in this section is based on the predictions obtained on the training data, as the core loss data of the test dataset is unknown. Table 2 lists the relative prediction errors of each material. It is noticeable that the prediction accuracy on the training dataset can only be used as a reference of model performance and do not fully reflect the generalizability of the model. The final judgment on model performance needs to be based on the model's performance on the test dataset.

C. Model interpretability

The proposed model obtains the core loss by simulating the dynamic hysteresis process and Eq. (1), rather than a simple numerical fitting like the Steinmetz formula. Fig. 6 shows an example of the predicted B-H curve for material B when the excitation is downsampled to 128 points.

Secondly, based on the behavior of magnetic materials under different magnetic field conditions, the model characterizes static hysteresis and dynamic hysteresis components respectively (see Eq. (2)). When capturing the static hysteresis component, the model considers two key magnetization mechanisms (domain rotation and domain wall motion) that cause magnetic moment changes from the magnetic domain level and reconstructs it using the stop hysteresis operator (see Eq. (3)). For the dynamic component, the model mainly focuses on field maldistribution and eddy current effects. Through the method of section segmentation and the powerful learning ability of neural networks, the calculation formula of the dynamic component under quasi-static conditions (see Eq. (4)) is transferred to high-frequency applications. As all modeling formulations rely on the principles of physics (Maxwell's Equations) and phenomenology, the authors are confident that

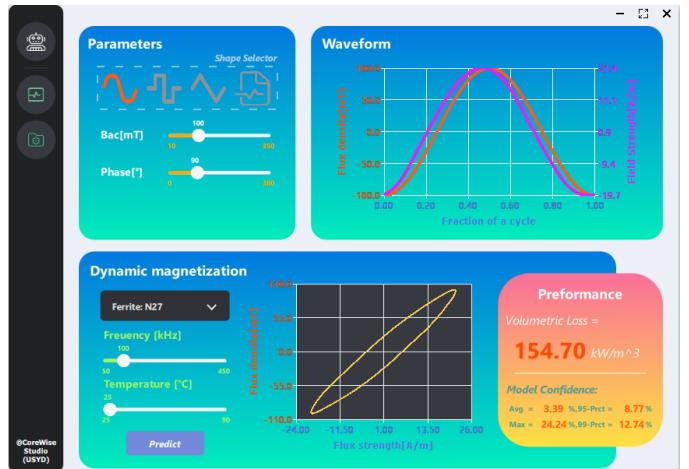


Fig. 7. Main page of the developed UI.

the proposed model possesses adequate physical insights to effectively represent the magnetization process and compute core losses of magnetic material.

D. Data visualization – QT based user interface

The user interface makes it easy to process and interpret data, ensuring a friendly experience for all users. We developed a UI based on QT design studio (see Fig.7), which contains the following features.

- Sine wave, trapezoidal wave, triangle wave and customized excitation inputs.
- B-H curve prediction and core loss prediction of a single cycle waveform.
- Core loss prediction of a dataset, and prediction data saved as csv file.
- Model accuracy evaluation through the histogram.

V. CONCLUSION

This paper introduces magnetization mechanism-inspired neural networks, a novel hybrid model capable of encoding underlying loss mechanisms and describing the dynamic magnetization process. More specifically, the proposed model consists of two subnetworks that represent the static and dynamic hysteresis components, respectively, based on the loss separation theory. The concepts of the PI hysteresis model and core loss ladder network segmentation are embedded in the subnetworks to enrich physics insights into the model. The MMINN applies a standard MSE loss function and an optimization algorithm (Adam) to train model parameters. In addition, a UI is designed to make data interpretation a visually engaging experience and help users explore B-H correlations.

Simulations of five materials were performed to verify the effectiveness of the proposed model, where the average 95% percentile error of the core loss predictions for the five materials is 5.1%. Besides, the presented networks have strong physical interpretability and compact size. Such outstanding model performance makes the proposed model suitable for matching with electromagnetic field simulation techniques (e.g., FEM) to drive advances in power magnetics automatic design.

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, doi: 10.1109/20.278656.
- [2] Willard, Matthew, "Stronger, Lighter, and More Energy Efficient: Challenges of Magnetic Material Development for Vehicle Electrification." 2013.
- [3] C. P. Steinmetz, "On the Law of Hysteresis," in Transactions of the American Institute of Electrical Engineers, vol. IX, no. 1, pp. 1-64, Jan. 1892, doi: 10.1109/T-AIEE.1892.5570437.
- [4] 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, July-Aug. 2001, doi: 10.1109/28.936396.
- [5] K. Venkatachalam, C. R. Sullivan, T. Abdallah and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," 2002 IEEE Workshop on Computers in Power Electronics, 2002. Proceedings., Mayaguez, PR, USA, 2002, pp. 36-41, doi: 10.1109/CIEPE.2002.1196712.
- [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, doi: 10.1109/TPEL.2011.2162252..
- [7] Jiali Li, T. Abdallah and C. R. Sullivan, "Improved calculation of core loss with nonsinusoidal waveforms," Conference Record of the 2001 IEEE Industry Applications Conference. 36th IAS Annual Meeting (Cat. No.01CH37248), Chicago, IL, USA, 2001, pp. 2203-2210 vol.4, doi: 10.1109/IAS.2001.955931.
- [8] Krings, Andreas & Soulard, Juliette , "Overview and Comparison of Iron Loss Models for Electrical Machines." Journal of Electrical Engineering. 10. 162-169, 2010.
- [9] Preisach, F. "Über die magnetische Nachwirkung. Z." Physik 94, 277-302, 1935.
- [10] Jiles, D.C. and Atherton, D.L. "Theory of ferromagnetic hysteresis." Journal of Magnetism and Magnetic Materials, Vol. 61, pp. 48-60, 1986. [https://doi.org/10.1016/0304-8853\(86\)90066-1](https://doi.org/10.1016/0304-8853(86)90066-1).
- [11] I. D. Mayergoyz, "Dynamic Preisach models of hysteresis," in IEEE Transactions on Magnetics, vol. 24, no. 6, pp. 2925-2927, Nov. 1988, doi: 10.1109/20.92290.
- [12] L. R. Dupre, O. Bottauscio, M. Chiampi, M. Repetto and J. A. A. Melkebeek, "Modeling of electromagnetic phenomena in soft magnetic materials under unidirectional time periodic flux excitations," in IEEE Transactions on Magnetics, vol. 35, no. 5, pp. 4171-4184, Sept. 1999, doi: 10.1109/20.799065.
- [13] T. Chevalier, A. Kedous-Lebouc, B. Cornut, C. Cester. "A new dynamic hysteresis model for electrical steel sheet." Physica B 275 (1-3) 197-201, 2000.
- [14] Y. Li, L. Zhu and J. Zhu, "Core Loss Calculation Based on Finite-Element Method With Jiles-Atherton Dynamic Hysteresis Model," in IEEE Transactions on Magnetics, vol. 54, no. 3, pp. 1-5, March 2018, Art no. 1300105, doi: 10.1109/TMAG.2017.2765704.
- [15] C. Nussbaum, H. Pfutzner, T. Booth, N. Baumgartinger, A. Ilo and M. Clabian, "Neural networks for the prediction of magnetic transformer core characteristics," in IEEE Transactions on Magnetics, vol. 36, no. 1, pp. 313-329, Jan. 2000, doi: 10.1109/20.822542.
- [16] M. Tian, H. Li and H. Zhang, "Neural Network Model for Magnetization Characteristics of Ferromagnetic Materials," in IEEE Access, vol. 9, pp. 71236-71243, 2021, doi: 10.1109/ACCESS.2021.3078554.
- [17] H. Li et al., "MagNet: An Open-Source Database for Data-Driven Magnetic Core Loss Modeling," 2022 IEEE Applied Power Electronics Conference and Exposition (APEC), Houston, TX, USA, 2022, pp. 588-595, doi: 10.1109/APEC43599.2022.9773372.
- [18] S. Quondam Antonio, Francesco Riganti Fulginei, Laudani, A., Gabriele Maria Lozito, & Riccardo Scorratti. "Deep neural networks for the efficient simulation of macro-scale hysteresis processes with generic excitation waveforms." Engineering Applications of Artificial Intelligence, 121, 105940-105940, 2023.
- [19] D. Serrano et al., "Why MagNet: Quantifying the Complexity of Modeling Power Magnetic Material Characteristics," in IEEE Transactions on Power Electronics, doi: 10.1109/TPEL.2023.3291084.
- [20] Raissi, M., Perdikaris, P., & Karniadakis, G. E. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." Journal of Computational Physics, 378, 686-707, 2019, <https://doi.org/10.1016/j.jcp.2018.10.045>.
- [21] Y. Wang, R. Xu and M. Zhou, "Prandtl-Ishlinskii Modeling for Giant Magnetostrictive Actuator Based on Internal Time-Delay Recurrent Neural Network," in IEEE Transactions on Magnetics, vol. 54, no. 11, pp. 1-4, Nov. 2018, Art no. 7301004, doi: 10.1109/TMAG.2018.2827397.
- [22] Grech C, Buzio M, Pentella M, Sammut N. "Dynamic Ferromagnetic Hysteresis Modelling Using a Preisach-Recurrent Neural Network Model." Materials. 2020; 13(11):2561, 2021, <https://doi.org/10.3390/ma13112561>.
- [23] G. Bertotti, "General properties of power losses in soft ferromagnetic materials," in IEEE Transactions on Magnetics, vol. 24, no. 1, pp. 621-630, Jan. 1988, doi: 10.1109/20.43994.
- [24] Bertotti, G. "Hysteresis in Magnetism." Academic Press: San Diego, CA, USA; pp. 391-430, 1998.
- [25] Zentková, A., Zentko, A. & Hajko, V. "The origin of negative Barkhausen jumps as a consequence of eddy currents in ferromagnets." Czech J Phys 19, 650-656, 1969.
- [26] Brokate, M., Sprekels, J. "Hysteresis Operators." in Hysteresis and Phase Transitions. Applied Mathematical Sciences, vol 121. Springer, New York, NY, 1996.
- [27] Mayergoyz, I. D. "Mathematical Models of Hysteresis and their Applications." Academic Press, 2003.
- [28] Prandtl, L. Ein Gedankenmodell zur kinetischen.. Theorie der festen Körper. ZAMM J. Appl. Math. Mech./Z. Angew. Math. Mech. 8, 85-106, 1928.
- [29] Everett, D.H., & Whitton, W.I. "A general approach to hysteresis." Transactions of The Faraday Society, 48, 749-757, 1952...
- [30] Marouani, H., Hergli, K., Dhahri, H. et al. "Implementation and Identification of Preisach Parameters: Comparison Between Genetic Algorithm, Particle Swarm Optimization, and Levenberg–Marquardt Algorithm." Arab J Sci Eng 44, 6941–6949, 2019...
- [31] L. Tian, "Two calculation methods of eddy-current losses in high frequency transformer," 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 2017, pp. 1-6, doi: 10.1109/CICT.2017.7977342.
- [32] S. Lin, T. Brinker, L. Fauth and J. Friebel, "Review of Dimensional Resonance Effect for High Frequency Magnetic Components," 2019 21st European Conference on Power Electronics and Applications (EPE '19 ECCE Europe), Genova, Italy, 2019, pp. P.1-P.10, doi: 10.23919/EPE.2019.8915152.
- [33] R.H. Pry and C.P. Bean. "Calculation of the energy loss in magnetic sheet materials using a domain model." J. Appl. Phys., Vol.29, No.3, pp532-533,1958.
- [34] Ionel, Dan & Popescu, Mircea & McGilp, Malcolm & Miller, T.J.E. & Dellinger, Stephen & Heideman, Robert. "Computation of Core Losses in Electrical Machines Using Improved Models for Laminated Steel." Industry Applications, IEEE Transactions, 2007, on. 43. 1554 - 1564. 10.1109/TIA.2007.908159.
- [35] Guyomar, D., Ducharme, B., & Sebald, G. "Dynamical hysteresis model of ferroelectric ceramics under electric field using fractional derivatives." Journal of Physics D, 40(19), 6048–6054, 2007.
- [36] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980, 2014.
- [37] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. "Optuna: a Next-Generation Hyperparameter Optimization Framework." in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

Prediction of Volumetric Loss of any Magnetic Material using Deep Neural Network - MagNet

Dixant Bikal Sapkota, Puskar Neupane, Mecon Joshi, Shahabuddin Khan

Abstract—This paper outlines the methodology for predicting power loss in magnetic materials. It starts by introducing the concept of core loss and the complexity of modeling it. Steinmetz's Equation is presented to calculate power loss based on frequency and magnetic flux density, but its limitations are highlighted. As an alternative, neural network based method is introduced. The proposed methodology adopts a Long Short-Term Memory Network, expressing the core loss as a function of magnetic flux density (B), frequency (f), temperature (T) and wave classification. Fast Fourier Transform was implemented to reduce the data points of the sampled B waveform while preserving its wave characteristics. The analysis in frequency domain enabled to streamline the training of the dataset. The data set was arranged as required, and the network architecture was designed with appropriate layers and optimal activation functions. Through training on the dataset, the network assimilated the intricate relationships between input variables and known power loss. Evaluation and validation metrics were subsequently employed to gauge the performance of the trained network. This innovative methodology aims to improve the precision of power loss predictions, providing valuable insights into the nuanced behavior of magnetic materials.

Index Terms—MagNet, Frequency Domain, LSTM, FFT

I. INTRODUCTION

MAGNETIC materials are typically characterized by their magnetic domain existence. In the absence of an external magnetizing field (H), these domains are aligned randomly. However, in the presence of such a field, these domains adjust themselves to align in parallel with the direction of the applied field. When an alternating current is applied, the domain's alignment oscillates with the applied field frequency, causing core loss, a phenomenon where heat is generated due to repeated changes in direction and magnitude of the domains.

Charles Steinmetz first developed the equation to characterize the magnetic core loss in 1892. It is an empirical formula that is used to compute the overall losses of power in any magnetic material expressed as power loss per unit volume. It was later modified to include frequency term and named as Steinmetz's Equation [1]. The equation is mathematically expressed as:

$$P_v = kf^a B^b \quad (1)$$

Here, P_v is the average power loss per unit volume, f is the frequency, B is the peak value of magnetic flux density and k , a , and b are Steinmetz constants whose values are obtained through B-H curve fitting of the given material. This equation gives a simple understanding of core losses and is a decent equation for general design of equipment. But in the field of power electronics, it is perhaps the weakest link due to

its ignorance of the impact of various other factors such as waveform shape, DC bias, ambient and core temperatures, etc. on power losses.

Over the years following, various models were developed by improving upon the Steinmetz equation. For instance, MSE, GSE and iGSE improved the results under non-sinusoidal excitation. Particularly, the improved generalized Steinmetz Equation (iGSE) has been considered an excellent calculation method. However, it gives large error for very high or very low duty cycles [2]. Similarly, other methods such as EEL and WcSE give high errors when duty cycle is far from 0.5 and methods such as DNSE, i²GSE, RESE and FGSE introduce new parameters which are not typically available in data sheets [2]. This shows the complexity of the process of modeling magnetic core loss.

In recent times, models based on neural networks are being considered for magnetic core loss prediction due to their effectiveness in solving non-linear problems. These models are data based in nature and can predict losses for any given material subjected to a certain field density with known frequency, temperature and additional related features. This approach requires careful study of impacting parameters, design of appropriate model architecture, large and accurate data set and optimal training process. This is the approach taken in this paper. Utilizing the power of the Python programming language and a large number of available libraries such as *numpy*, *pandas*, *tensorflow*, and *keras*, a model has been designed that can predict losses for any magnetic material with decent accuracy and a relatively small number of input features.

II. PRELIMINARY WORKS/LITERATURE REVIEW

Previous studies in this field have made essential datasets available for the model's training, validation, and testing [3], [4], [5], significantly contributing to the preparatory phases of this study. These datasets assume a pivotal role in refining and evaluating the model.

[6] surveys various publications, exploring the application of AI in Power Electronics to enhance accuracy and efficiency in the sector. In Paper [7], the datasheets for loss and B-H loop utilizing neural networks are shown, along with an illustration of how beneficial neural networks are in forecasting core losses under various operating conditions. The DNN method was implemented in Paper [8], which accurately modeled the core losses and recommended capturing additional scenarios to enhance the model. The most advanced neural network models have also been studied, and they have shown to be more accurate in predicting core losses than traditional machine learning

models. The transformer neural network model described in Paper [9] has proven to be effective in simulating hysteresis loops by symbolically integrating sequence inputs and scalar inputs.

III. FOURIER TRANSFORM AND FFT

Fourier Transform [10] is a mathematical technique which decomposes any signal which is a function of time into its various constituent frequency components. This is a transformation between time-domain representation and frequency-domain representation.

Fast Fourier Transform, abbreviated as FFT, is a mathematical algorithm that computes the Discrete Fourier Transform [11]. The "numpy" library in python provides functionality for calculating one dimensional n-point discrete fourier transform with the FFT algorithm. The input to the FFT algorithm is an array of sequential data points representing a sampled time series data and its result is an array of complex numbers representing the samples of frequency domain spectrum. The Fast Fourier Transform (FFT) has been used here to decrease the computational complexity of the model by reducing the input parameters required for calculation.

IV. METHODOLOGY

A. Data Processing

The provided training dataset consists of five different comma separated values (csv) files for every material, each containing data for magnetic field density (B), applied magnetizing field force (H), frequency (F), temperature (T) and measured volumetric loss (V). The number of rows of data is not same for all materials and is listed in Table I. There are 1024 time instance values for B and H in a single cycle and a single value for others. Additional data was provided for the testing of the trained model.

TABLE I
TOTAL NUMBER OF PROVIDED DATA FOR EACH MATERIALS

Material	Training Data	Testing data	Type
3C90	40713		
3C94	40068		
3E6	6996		
3F4	6564		
77	11444		
78	11380	5000	Known
N27	11396		
N30	8978		
N49	8602		
N87	40616		
A	2432	7651	
B	7400	3172	
C	5357	5357	Unknown
D	580	7299	
E	2013	3738	

To reduce the computational time for training the models, there was a need to minimize the data size while preserving as much valuable information as possible to sustain accuracy. To achieve this, the focus was primarily set on the aggregation of the given waveform data. Various methods were explored,

including selecting only a few of the numerous provided time instance values, averaging the values over a specific range, and decomposing the wave into its constituent harmonics but the latter seemed to give the most favorable results with higher consistency.

Fast Fourier Transform, shortened as FFT, was employed to transform each of the provided time domain waveform into its frequency domain representation. Specifically, the DC offset and the next 31 harmonics (i.e. first 32 terms in total) were used for model training. This technique preserved the wave characteristics during the training while also reducing the overall data input directly to the model. Since the output of FFT is a complex quantity, its absolute value was considered for each harmonic. This was done because if both real and imaginary values are to be used for model training, there would be twice as much data, which would increase the computational needs. The model's accuracy was observed to be largely unaffected by the inclusion of more harmonics; so, the terms were restricted to 32.

Additionally, further techniques were integrated to efficiently recognize and classify various waveforms. The waves were categorized into sinusoidal, triangular and trapezoidal waves based on their frequency domain representations. Sine waves have high fundamental component and negligible harmonics thereafter, while triangular waves have low fundamental component to second harmonic ratio and trapezoidal waves have high value of the same ratio. These properties were used for the classification and it seemed to give an acceptable classification although there were some outliers in trapezoidal and triangular category. Practically, sinusoidal waves were isolated first, and then waves having fundamental to second harmonic ratio less than 7 were classified as triangular while more than that were classified as trapezoidal. An example of this can be seen in Figure 1.

The goal is to establish a cost-effective equilibrium between model accuracy and resource efficiency through the reduction of input data size and the proper classification of waveforms. After these calculations, the data was re-arranged in a separate array with the labels as shown in Table II.

TABLE II
DATA ARRANGEMENT LABELS

S.N	B1, B2, B3,... up-to B32	Freq	Temp	Loss	Classification
-----	--------------------------	------	------	------	----------------

B. Data Visualization

In order to fully understand the nature of the parameters for each material, various plots were drawn. A scatter plot drawn between different features for 3C90 can be seen in Figure 2. It can be observed that the data doesn't have any defined correlation. High volumetric losses exist for lower B_{max} , but even at higher fields volumetric loss ranges significantly due to changes in frequency and temperature values. From Figure 2(d) it can be observed that high frequencies are only present for lower field densities. This is proved by the correlation between B_{max} and Frequency being -0.430108. The effect of this can be observed in Figure 2(b) where, as we proceed

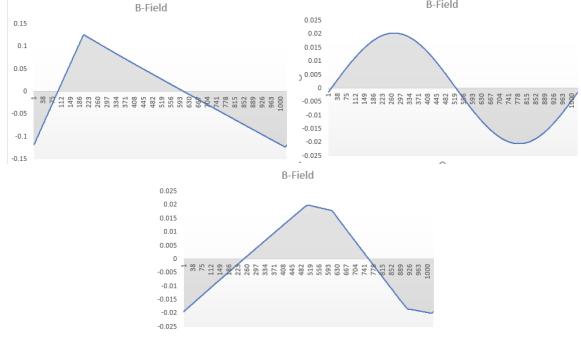


Fig. 1. Classification of magnetic field density waveforms: 0 for triangular, 1 for sinusoidal and 2 for trapezoidal

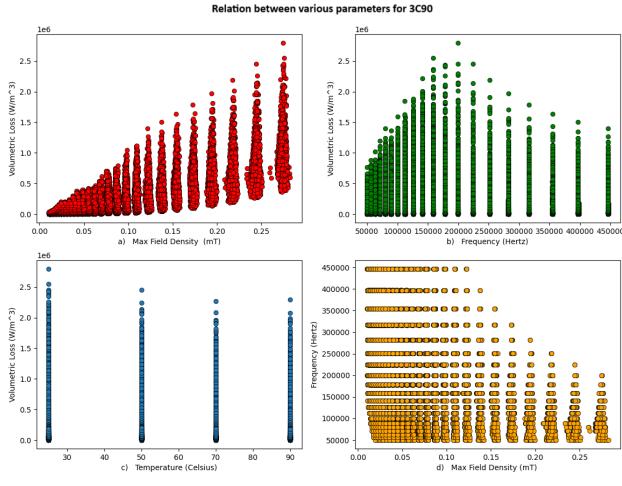


Fig. 2. Scatter plot between various parameters for Material 3C90: a) Maximum Field Density vs Volumetric Loss, b) Frequency vs Volumetric Loss, c) Temperature vs Volumetric Loss, d) Maximum Field Density vs Frequency

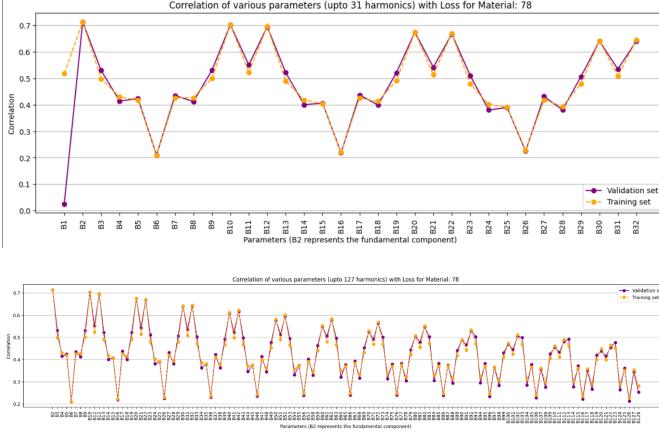


Fig. 3. Correlation plot of Loss vs various parameters for Material 78 with (a) 31 Harmonics and (b) 127 Harmonics

to higher frequencies, the overall volumetric losses slightly decrease. Using `pandas.DataFrame.corr()`, the correlation between B_{\max} and Volumetric Loss was found to be 0.759792. In the frequency domain, Figure 3 shows the correlation between various harmonics and loss. Here, B2 represents

fundamental or first harmonic, B3 represents second harmonic and so on. It can be seen that different harmonics have different correlations. From Figure 3(a) it can be seen that the 1st, 9th, 11th, 19th, 21st, 29th, 31st harmonics seem to have a much higher level of correlation and this pattern continues in a descending gradient as seen in Figure 3(b).

C. Model Architecture

The neural network model is a multi-input architecture, strategically tailored to address a specific task. There are four different input layers in the model, and each has a specific function. The first input layer is capable of processing data sequences having a particular shape of (32, 1) since it is set up to handle the waveform information. This arrangement corresponds to the expectation of input sequences with length of 32. To extract any temporal dependencies in this frequency domain, two Long Short-Term Memory (LSTM) layers with 128 units each were employed as seen in Figure 4. These LSTM layers play a pivotal role in capturing intricate patterns within the input. To augment the stability of the training process, batch normalization and dropout mechanisms are incorporated in between these two LSTM layers.

TABLE III
DENSE LAYER INFORMATION

Layer	Type	Neurons	Connected to
A	Dense	1024	LSTM Out
B	Dense	512	A
C	Dense	256	B
D	Dense	128	C
E	Dense	64	D
F	Dense	32	E
G	Dense	16	F
Output	Dense	1	G

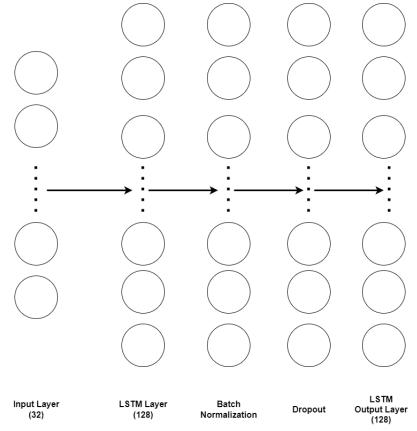


Fig. 4. LSTM Layer Diagram

The output features obtained from the LSTM layers are subsequently concatenated with the remaining inputs which are frequency, temperature, and classification. This consolidated input is then fed into the subsequent layers of the neural network. The processing of this concatenated input is executed through several dense layers, each equipped with Rectified

Linear Units (ReLU) as activation functions. ReLU activation is a piece-wise linear function which signifies that the node gets activated only if the sum of weighed inputs is greater than 0, else, the node is deactivated, i.e. turned off. ReLU is used here because of its simplicity, general use case and for reduction of vanishing gradient. The information of each of the dense layers is tabulated in Table III.

TABLE IV
MODEL TOTAL PARAMETERS

Material	Model Layer	Total Params	File Size
A	LSTM+ Dense	1033729	3.94 MB
B	LSTM+ Dense	1033729	3.94 MB
C	LSTM+ Dense	1033729	3.94 MB
D	LSTM+ Dense (C to G)	276225	1.09 MB
E	LSTM+ Dense	1033729	3.94 MB

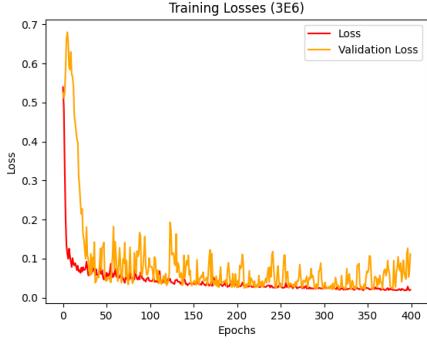


Fig. 5. Training and Validation loss for Material: 3E6

D. Training

Each model was trained for 500+ epochs. During training, *ModelCheckpoint* under *keras.callbacks* saves the model after a complete epoch if, at that point, it has a lower value of *validation loss*. The plot of training and validation losses, as seen in Figure 5 for the material 3E6, shows that the model learns at a very high rate within the first few epochs, and gradually saturates over time. The noisy nature of Validation loss is due to a smaller batch size which allows the model to learn from each individual example of the training set.

V. RESULTS

A. Error Calculation

Error for each instance is calculated using Equation 2

$$\text{Error}(\%) = \frac{|meas - pred|}{meas} * 100\% \quad (2)$$

Here, *meas* represents the measured loss provided in the dataset, *pred* represents the prediction of loss outputted by the model. The predictions are outputted in a new list whose values range within the scaling limits. These values are all transformed back using *inverse_scaler* to their original range before the error is calculated.

B. Histograms

To analyze the performance of the models, histogram is plotted with percentage relative error as the x-axis and ratio of data points as the y-axis for each material, using *matplotlib*. The plot highlights and labels the overall average error, 95 percentile error and the maximum error. For the known materials, error distribution histogram was plotted for a different validation data-set that was provided separately. However, for the unknown materials, the histogram was plotted for a testing set extracted from the training data using the *train_test_split* function.

C. Testing Results

After training the model using provided data-set for each of the 10 different materials, each model's validity was individually tested using separate validation data sets, each consisting of 5000 randomly sampled data from the original database. The performance of the model for each material can be observed in the histograms as seen in Figures 6. The model was re-fit and tested on 5 unknown materials - A, B, C, D and E. Their validation error histograms can be seen from Figure 7. These sets were provided as a part of the final evaluation of the performance of the model. The training set here varied from thousands of data points for some to a few hundreds of data points for others. In some cases the model was entirely re-trained, whereas in some, a favorable model was re-fit with these additionally provided data points.

VI. MODEL INFORMATION

Table V lists the model details. Each material has its own separate model. The model architecture used for each material is the same, however, certain parameters like epochs, batch size, learning rate etc. are varied for optimal results.

TABLE V
MODEL STATISTICS

No. of input features to model	35
H5 model file size	approx 3 MB
Scaling file size (.joblib file)	1 KB
No. of csv files required before execution	3

VII. CONCLUSION

This research presents a neural-network based approach to predict magnetic core loss. To reduce training complexity, the waveforms are converted to frequency domain by decomposition into harmonics while simultaneously analyzing the waveforms to classify them. The designed neural-network consists of several LSTM layers for feature extraction followed by dense layers with ReLU activation. The model was found to be accurate in predicting core loss for the known materials as shown by the error histograms. Also, the methodology showed promising results in the training data of unknown materials.

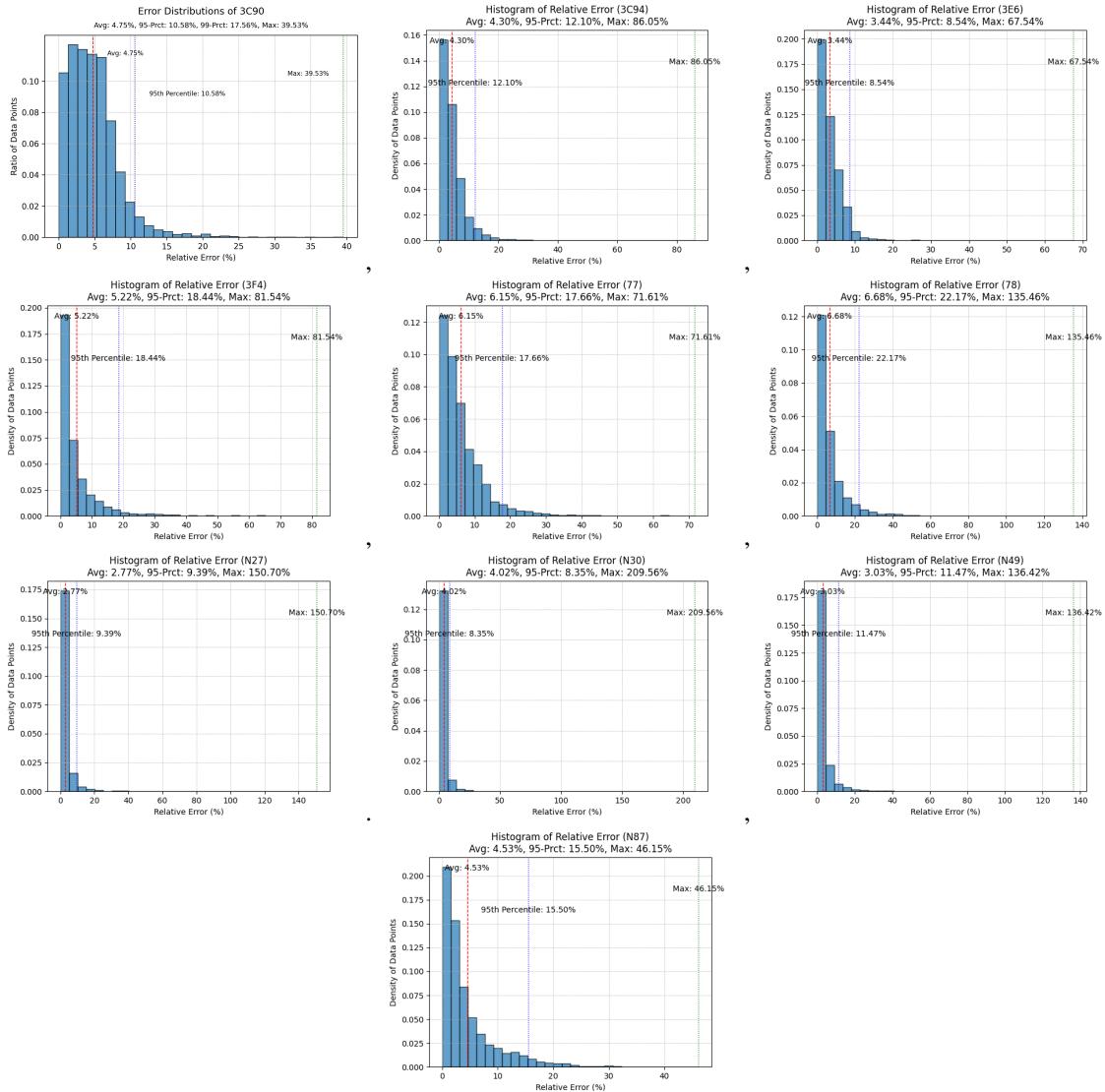


Fig. 6. Error Histogram of 10 different known materials

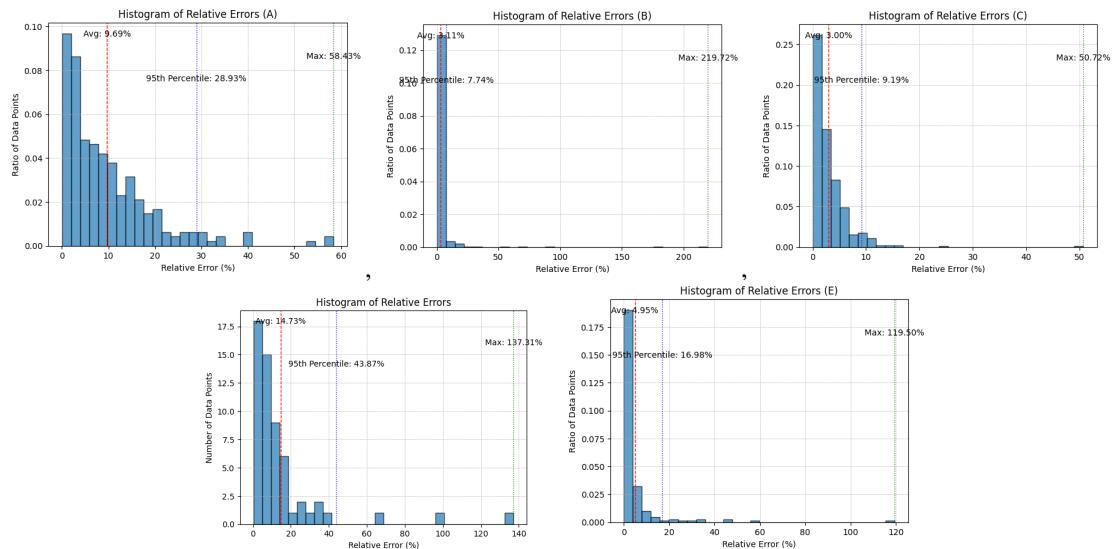


Fig. 7. Validation Error Histogram of 5 different unknown materials

REFERENCES

- [1] E. Snelling, *Soft Ferrites Properties and Applications*. Butterworth amp; Co, 1988.
- [2] Z. Li, W. Han, Z. Xin, Q. Liu, J. Chen, and P. C. Loh, "A review of magnetic core materials, core loss modeling and measurements in high-power high-frequency transformers," *CPSS Transactions on Power Electronics and Applications*, vol. 7, no. 4, pp. 359–373, 2022.
- [3] 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," *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14 292–14 316, 2023.
- [4] H. Li, D. Serrano, T. Guillod, S. Wang, E. Dogariu, A. Nadler, M. Luo, V. Bansal, N. K. Jha, Y. Chen, C. R. Sullivan, and M. Chen, "How magnet: Machine learning framework for modeling power magnetic material characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15 829–15 853, 2023.
- [5] H. Li, D. Serrano, S. Wang, and M. Chen, "Magnet-ai: Neural network as datasheet for magnetics modeling and material recommendation," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15 854–15 869, 2023.
- [6] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," *IEEE Transactions on Power Electronics*, vol. 36, no. 4, pp. 4633–4658, 2021.
- [7] D. Serrano, H. Li, T. Guillod, S. Wang, M. Luo, C. R. Sullivan, and M. Chen, "Neural network as datasheet: Modeling b-h loops of power magnetics with sequence-to-sequence lstm encoder-decoder architecture," in *2022 IEEE 23rd Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2022, pp. 1–8.
- [8] 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)*, 2022, pp. 1–8.
- [9] H. Li, D. Serrano, S. Wang, T. Guillod, M. Luo, and M. Chen, "Predicting the b-h loops of power magnetics with transformer-based encoder-projector-decoder neural network architecture," in *2023 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2023, pp. 1543–1550.
- [10] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.
- [11] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70–73, 1967. [Online]. Available: <https://api.semanticscholar.org/CorpusID:13900622>

MagNet Challenge 2023 Final Report

Bowen Su, Yunhao Xiao, Min Yang, *Student Member, IEEE*, Kai Sun, *Fellow, IEEE*

Abstract- The power loss in magnetic materials when subjected to external changing magnetic flux is very important and can affects the efficiency of power equipment. However, most of the computing methods for losses in magnetic materials are still empirical fitting methods based on Steinmetz's equation. And the accuracy is affected by the Steinmetz's coefficients which vary with working conditions. A more accurate and flexible method is needed for the computation of losses in magnetic materials. Based on the database provided by the MagNet competition, this paper firstly introduces a loss prediction model using multilayer perceptron, achieving proper prediction results on different materials. Furthermore, to address the limitations of manually extracting input data features, a new prediction model for loss of magnetic materials is proposed based on the attention mechanism, further enhancing the accuracy of magnetic material loss predictions. The maximum and minimum relative errors for N87 are 24.71% and 0.003% respectively.

I. INTRODUCTION

POWER losses are important performance parameters in the power equipment. It not only affects efficiency but is also closely related to the temperature of power equipment. Generally speaking, power losses can be divided into copper loss and iron loss. Copper loss refers to the power loss generated by the conductor when current flows through it. Currently, there have been many theoretical and computational methods available, such as the Dowell model and Tourkhani model [1], [2]. Iron loss refers to the power loss generated by magnetic materials when subjected to external changing magnetic flux. However, due to the lack of clear underlying physical models, there are rarely theories and computational models available for computing iron loss.

Currently, the main computation methods for iron loss are the Steinmetz's equation and its variations, including the Modified Steinmetz Equation, the Generalized Steinmetz Equation and Improved Steinmetz Equation [3]-[6]. These models are essentially based on the Steinmetz model, and their accuracy relies on the coefficients k , α , and β . However, the coefficients are very sensitive to operating conditions such as frequency and temperature, leading to significant errors in the existing methods. For example, the fitting coefficients α and β in the Steinmetz model will be affected by conditions such as different frequencies and magnetic induction intensities [7]. While current Steinmetz models and its variants takes the fitting coefficients α and β as constants. In addition, the above models hardly take into account the influence of temperature on the magnetic material losses. And the differences in losses at different temperatures can be

significant.

To address the above issues, there are two possible improvement approaches. On the one hand, a completely new model for calculating iron losses can be developed by analyzing and summarizing the characteristics of measured data sets. On the other hand, modern technologies such as artificial intelligence can be used to fit the Steinmetz coefficients using different methods, aiming to improve the accuracy of the formulas as much as possible. Numerous studies have been proposed in the field and are working towards the modeling of data-driven models for loss of magnetic material. A Machine Learning (ML)-based approach of identification for hysteresis and its inverse model are proposed in [8]. In [9], a neural network hysteresis model for the prediction of magnetic properties of ferromagnetic materials based on BP neural network is introduced, which only considered the influence of frequency on the magnetic properties. Recently, three neural network frameworks of the loss model of magnetic materials which cover different kinds of input and output, as well as a loss measurement system which can acquire large-scale and high-quality database are introduced in [10]. Furthermore, a transformer-based encoder-projector-decoder neural network architecture for modeling power magnetics B–H hysteresis loops is proposed, which can predict the losses of magnetic materials [11], [12]. However, magnetic material loss predictions based on hysteresis loops do not take into account the actual eddy current losses. In [13], a Deep Neural Network (DNN) approach to core loss estimations is proposed, which only considered the loss of N87 under a sinusoidal excitation.

Benefiting from the high-quality dataset provided by MagNet [14], this paper proposes a novel neural network structure for magnetic materials loss prediction, achieving high accuracy and model adaptability which can be used for different magnetic materials. In section II, the preliminary multi-layer perceptron regressor model for loss prediction of magnetic materials based on multiple input features extraction is proposed. Furthermore, an approach based on attention mechanism is proposed, which can address the limitations associated with the manual selection of features in the preliminary model. Section III gives the results of prediction errors for the testing materials.

II. PREAMBLE DATA ANALYSIS AND AN OVERVIEW OF MULTI-LAYER PERCEPTRON MODEL

A. Data analysis

Firstly, the data provided on April 1st, 2023 includes single cycle magnetic flux density waveforms (1024 discrete points) of ten different materials under various excitations, as well as

Bowen Su, Min Yang, Yunhao Xiao and Kai Sun are with the Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: subw20@mails.tsinghua.edu.cn; yangmin20@mails.tsinghua.edu.cn; xiaoyh23@mails.tsinghua.edu.cn; sun-kai@mail.tsinghua.edu.cn).

the corresponding excitation waveform frequencies, operating temperatures, and single cycle magnetic field intensity waveforms (1024 discrete points).

After analysis, the provided single cycle magnetic flux density waveforms have significant differences, not only in the standard sine wave waveform, but also in a variety of magnetic flux density waveforms with different harmonic contents, as shown in Fig. 1.

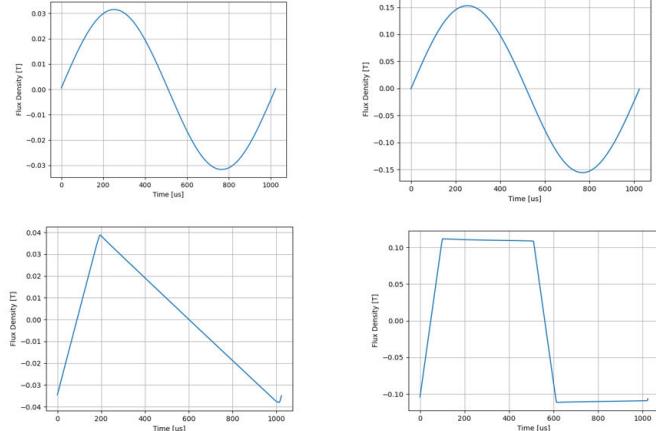


Fig.1 Waveforms of different magnetic flux density

Secondly, the data provided for final evaluation on November 10th, 2023 includes single cycle magnetic flux density waveforms (1024 discrete points) of 5 different materials under different excitation waveforms, as well as the corresponding frequency and operating temperature of the excitation waveforms. And consistent with the data provided for the first time, the provided single cycle magnetic flux density waveform still has significant differences, containing multiple magnetic flux density waveforms with different harmonic contents, as shown in Fig. 2.

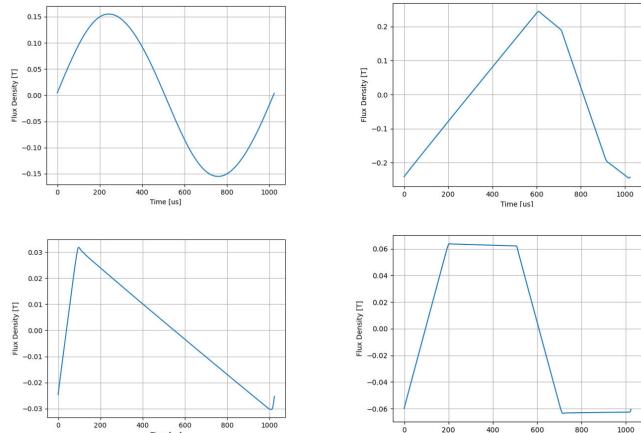


Fig.2 Waveforms of different magnetic flux density in final evaluation

However, it can be observed that the waveform shape of the data provided for testing is roughly the same as the waveform provided for the first time. Therefore, the deep learning model trained from the waveform provided for the first time can be well transferred to the waveform used for testing, generating robust magnetic core loss prediction results.

B. Multi-layer perceptron regressor model

According to the data analysis in the previous text, it can be seen that the given data mainly consists of magnetic flux density waveforms and related working conditions, and the magnetic flux density waveforms are relatively complex, containing various harmonics with different frequencies and amplitudes.

Therefore, the calculation of various features that comprehensively reflect the waveform properties of magnetic flux density is envisioned, followed by the determination of magnetic core loss density through the utilization of a Multi-layer Perceptron (MLP) regressor.

Concerning feature selection in the time-domain, the amplitude, peak-to-peak value, effective value, and variance of the magnetic flux density waveform are directly extracted as four dimensional features. Subsequently, two types of dimensionless features, namely skewness and kurtosis, of the magnetic flux density waveform are computed. The calculation formula is shown in (1), (2).

$$S = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma_i} \right)^3 \quad (1)$$

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma_i} \right)^4 \quad (2)$$

Regarding frequency domain characteristics, Fast Fourier Transform (FFT) is initially applied to the magnetic flux density waveform. Essential frequency domain features, including fundamental amplitude and frequency, second harmonic amplitude and frequency, as well as third harmonic amplitude and third harmonic frequency, are then extracted from the spectrum of the magnetic flux density waveform.

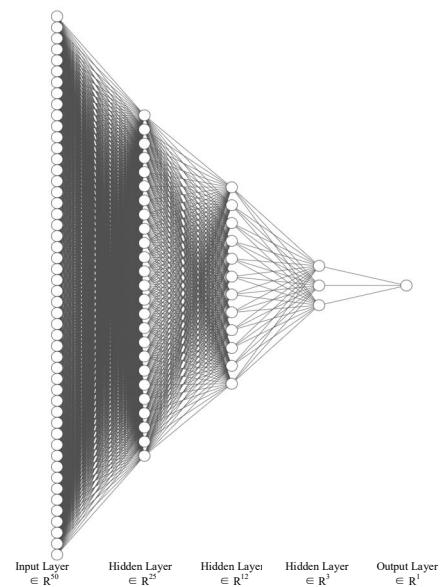


Fig.3 The structure of MLP

To effectively capture potential coupling relationships among the selected features, polynomial features of degree 3 or higher are incorporated for all features. Utilizing the product method, the relationships between the aforementioned features are indirectly characterized within the newly introduced polynomial features.

The MLP regressor consists of three fully connected layers, with each layer corresponding to a number of neurons (210, 150, 30). And after multiple tests, the GELU function is chosen as the activation function, which has higher accuracy and is more conducive to the back propagation process. The model structure is shown in Fig. 3.

However, this method of manually selecting features and applying MLP regressors for core loss prediction has two drawbacks. The first issue is that the model parameters are too large. In order to better preserve waveform information, more

features will be selected, which leads to an increase in the number of neurons in the fully connected layer, thereby greatly increasing the overall parameters of the model; The second reason is that the model accuracy is relatively low, as the manual feature selection process is greatly influenced by the operator's professional knowledge and experience, and the features obtained from different magnetic flux density waveforms through the same feature selection process may vary greatly, resulting in lower robustness.

III. THE PROPOSED ENCODER-BASED MODEL

In order to overcome the drawbacks of manually selecting features and applying MLP regressor methods mentioned above, an attention mechanism based approach from sequence to scale is proposed, with the structure shown in Fig. 4.

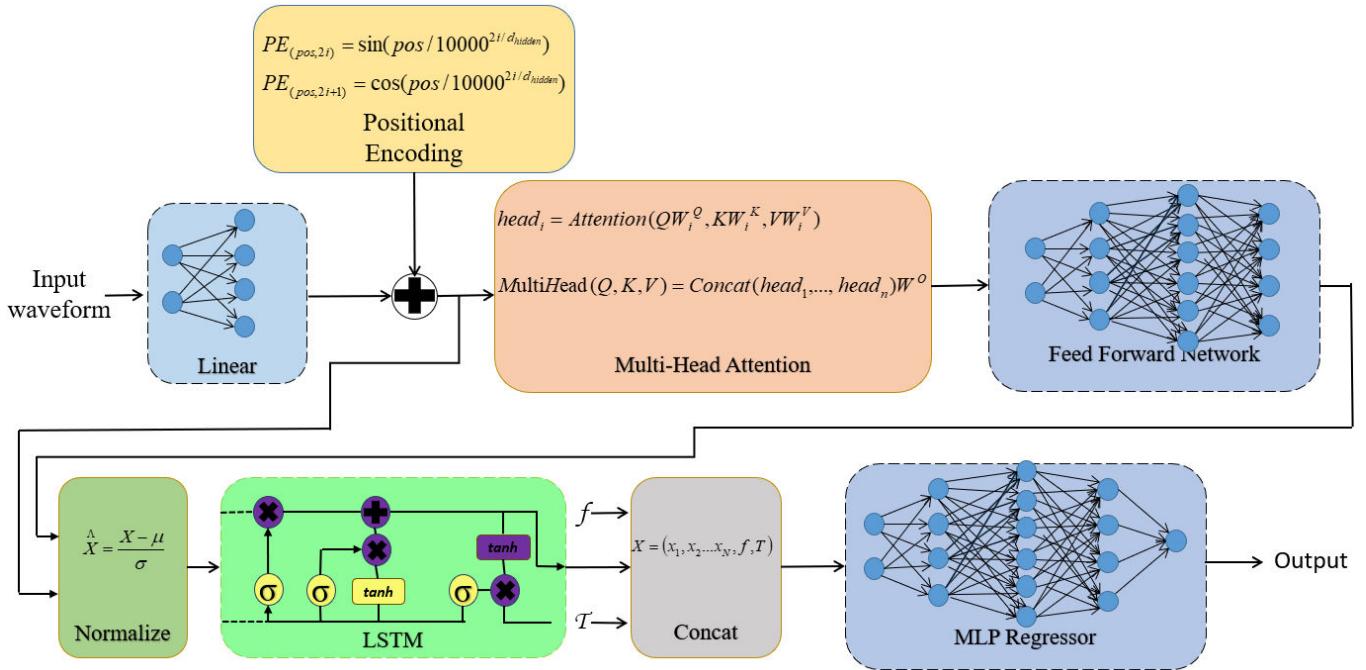


Fig.4 The structure of encoder-based model

In this method, the magnetic flux density waveform with 1024 discrete points is first directly taken as input, and each discrete single point value is expanded into a 64-dimensional vector. After expansion, more information may be stored during the training process, which is more conducive to subsequent attention calculation. Then, its position is encoded by (3), (4).

$$PE_{(pos,2i)} = \sin(pos / 10000^{2i/d_{hidden}}) \quad (3)$$

$$PE_{(pos,2i+1)} = \cos(pos / 10000^{2i/d_{hidden}}) \quad (4)$$

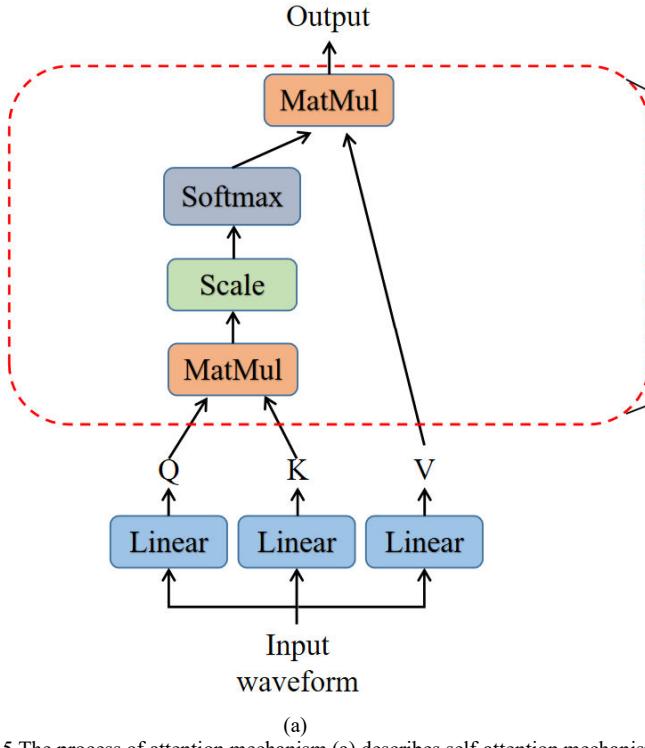
The use of sine function as position encoding mainly takes into account that due to the unique mathematical manipulation property of the sine function, any dimensional position encoding of any point can be linearly represented by the position encoding of other points, thus effectively

characterizing the relative position relationship between points.

Afterwards, the waveform with added position encoding will be input into an encoder layer[15] with self-attention mechanism. In this layer, the internal relationships of the magnetic flux density waveform will be constructed through self-attention mechanism, and the relationships between different points will be fully constructed through different attention scores, which is calculated as the formula (5). After passing through the encoder layer, each vector contains information about the entire waveform and has different values depending on the attention score. And the whole process of the attention mechanism is shown in Fig.5.

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (5)$$

The attention mechanism currently lacks a comprehensive and intuitive mathematical explanation, but intuitively speaking, it models the strength of the interrelationships between different input vectors by extracting the Query matrix and Key matrix of the input matrix and performing matrix dot multiplication on them. Intuitively speaking, the attention mechanism allows the model to see the key



(a)

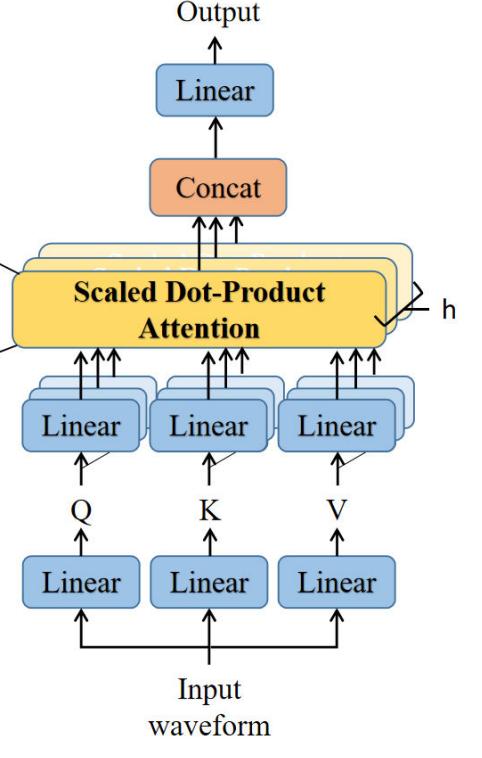
Fig.5 The process of attention mechanism.(a) describes self-attention mechanism and (b) describes the multi-head attention mechanism.

Then, feature extraction is performed using a Long Short-Term Memory (LSTM)[16] network. A matrix containing 1024 vectors is input into the LSTM layer, and each vector of the matrix is treated as a sequence of information. Through continuous iteration of the LSTM, the last intermediate state is finally taken as the enrichment of the entire matrix information.

Finally, the extracted information vector is partially compressed through a simple fully connected network, then concatenated with the excitation waveform frequency and operating temperature, and input as input features into the MLP regressor. A simple double-layer fully connected network with a structure of (16,8) is used to predict the magnetic core loss density.

Overall, the proposed model mainly utilizes the encoder layer and its inherent self-attention mechanism to expand the information and model the internal relationships of the magnetic flux density waveform, then uses LSTM for feature extraction, and finally uses MLP regression to predict the magnetic core loss density. In fact, the adoption of self-attention mechanism plays an excellent role in modeling the internal relationships of waveforms, and this mechanism does not bring excessive parameter increase or other negative

information and possible connection relationships in the input data, effectively modeling the internal relationships of the input waveform. In this case, after the input waveform matrix passes through an encoder layer containing self attention, each vector in the matrix contains the overall information of the matrix, which also provides the possibility for effective feature extraction in subsequent structures.



(b)

effects to the entire model. Moreover, since the self-attention mechanism is independent of parameters, it can be applied to different input waveforms, even if these waveforms may be very complex and very different from those in the training data, which undoubtedly greatly increases the robustness of the model and effectively improves the accuracy of the model.

Although the above model has good performance in accuracy and robustness, it also has a significant drawback, which is its high time complexity. Because time-series processing models such as Encoder layer and LSTM are used in the model, it means that the complexity of the model is closely related to the length of the processed sequence. For the self-attention mechanism, for a sequence of length N, its complexity is $O(N^2)$, while for LSTM, for a sequence of length N, its complexity is $O(N)$. Therefore, when the sequence length is long, that is, when the sampling rate of the single cycle magnetic flux density waveform is high, the computational complexity of the model will significantly increase. Correspondingly, the training time and deployed running time will also increase significantly.

In order to further improve accuracy, some hyperparameters can be adjusted and optimized, such as the number of layers in the encoder layer in this model. Using

multi-layer encoders can more accurately model the internal relationship of the magnetic flux density waveform, but it should be noted that this will increase the model parameters and computational complexity. In addition, the feature extraction part of the model can be modified. Except for LSTM, structures such as MLP or convolutional neural networks can also play a role in feature extraction. In summary, there is still room for further improvement in this model.

IV. RESULTS

The training process for the five materials used for final evaluation is basically similar. Firstly, the given training data for each material is divided into a training set and a validation set in a 7:3 ratio, and the hyperparameters are adjusted on the validation set.

And the Adam optimizer[17] is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\xi = 10^{-9}$. The decay of the learning rate over the course of training is as shown in (6).

$$lrate = lrate_{initial} \times (0.5)^{\text{epoch}/100} \quad (6)$$

This corresponds to decreasing the learning rate as the number of training epoch increases, which is beneficial for the model to converge to its optimal point. The number of parameters of each model for different material is shown in TABLE I.

TABLE I Number of parameters in models for different materials

Materials	A	B	C	D	E
Parameter-number	116061	116061	116061	116061	116061

And the test results of the two methods mentioned earlier, namely MLP regressor model and Encoder based model, on material N87, are shown in Table II.

TABLE II Test results of different models for N87 material

Method	Min. error	Avg. error	Max. error	95 th . error	R2 Score
MLP	0.002%	13.642%	101.964%	23.336%	0.974
LSTM	4.791%	3.668%	26.185%	9.439%	0.997
Proposed model	0.003%	3.568%	24.717%	8.707%	0.998

V. CONCLUSION

This paper utilizes the dataset provided by MagNet. Firstly, a magnetic material loss prediction model is proposed based on a feedforward neural network, focusing on the extraction of magnetic flux density waveform features. The testing errors for different materials are maintained within 10%. Considering the limitations of manually extracting waveform features, this paper further proposes an encoder-based model which contains attention mechanism to address the limitations associated with the manual selection of features. Compared with the MPL model and LSTM model, the proposed encoder-based model demonstrates higher accuracy in predicting the losses of magnetic materials.

In the future, there are still many potential directions for the development of proposed model. On the one hand, incorporating some physical mechanisms into the consideration of neural network models can enhance the interpretability of the neural network at the physical level. On the other hand, further adjustments to the network structure can be made to extract input features more effectively, thereby improving the model's accuracy.

REFERENCES

- [1] P. L. Dowell, "Effects of eddy current in transformer windings", *Proc. Inst. Electr. Eng.*, vol. 113, no. 8, pp. 1387-1394, Aug. 1966.
- [2] F. Tourkhani and P. Viarouge, "Accurate analytical model of winding losses in round Litz wire windings," in *IEEE Transactions on Magnetics*, vol. 37, no. 1, pp. 538-543, Jan 2001.
- [3] C. P. Steinmetz, "On the law of hysteresis," in *Proceedings of the IEEE*, vol. 72, no. 2, pp. 197-221, Feb. 1984.
- [4] 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, July-Aug. 2001.
- [5] Jieli Li, T. Abdallah and C. R. Sullivan, "Improved calculation of core losses with nonsinusoidal waveforms," *Conference Record of the 2001 IEEE Industry Applications Conference. 36th IAS Annual Meeting (Cat. No. 01CH37248)*, Chicago, IL, USA, 2001, pp. 2203-2210 vol.4.
- [6] K. Venkatachalam, C. R. Sullivan, T. Abdallah and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," *2002 IEEE Workshop on Computers in Power Electronics, 2002. Proceedings.*, Mayaguez, PR, USA, 2002, pp. 36-41.
- [7] P. Papamanolis, T. Guillod, F. Krismer and J. W. Kolar, "Minimum Loss Operation and Optimal Design of High-Frequency Inductors for Defined Core and Litz Wire," in *IEEE Open Journal of Power Electronics*, vol. 1, pp. 469-487, 2020.
- [8] X. Zhao and Y. Tan, "Modeling Hysteresis and Its Inverse Model Using Neural Networks Based on Expanded Input Space Method," in *IEEE Transactions on Control Systems Technology*, vol. 16, no. 3, pp. 484-490, May 2008.
- [9] M. Tian, H. Li and H. Zhang, "Neural Network Model for Magnetization Characteristics of Ferromagnetic Materials," in *IEEE Access*, vol. 9, p. 71236-71243, 2021.
- [10] H. Li *et al.*, "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," in *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15829-15853, Dec. 2023.
- [11] H. Li, D. Serrano, S. Wang, T. Guillod, M. Luo and M. Chen, "Predicting the B-H Loops of Power Magnetics with Transformer-based Encoder-Projector-Decoder Neural Network Architecture," *2023 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Orlando, FL, US

- A, 2023, pp. 1543-1550.
- [12] 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. 15854-1 5869, Dec. 2023.
- [13] X. Shen, H. Wouters and W. Martinez, "Deep Neural Network for Magnetic Core Loss Estimation using the MagNet Experimental Database," 2022 24th European Conference on Power Electronics and Applications (EPE'22 ECCE Europe), Hanover, Germany, 2022, pp. 1-8.
- [14] H. Li et al., "MagNet: An Open-Source Database for Data-Driven Magnetic Core Loss Modeling," 2022 IEEE Applied Power Electronics Conference and Exposition (APEC), Houston, TX, USA, 2022, pp. 588-595.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," arXiv:1706.03762 [cs.CL], Jun, 2017. DOI: 10.48550/arxiv.1706.03762.
- [16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997.
- [17] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs.LG], Dec, 2014. DOI: 10.48550/arXiv.1412.6980

MagNet Challenge 2023: Data-Driven Model for Power Loss Estimation of Magnetic Components - Final Report

Z. Li, *Student member, IEEE*, R. Mirzadaranji, *Student member, IEEE*, R. Liu, *Student member, IEEE*, L. Wang, *Student member, IEEE*, T. Luo, *Student member, IEEE*, D. Lyu, *Student member, IEEE*, M. Ghaffarian Niasar, Z. Qin, *Senior member, IEEE*

Abstract- Traditional methods like Steinmetz's equation (SE) and its improved version (iGSE) have shown limited accuracy in calculating power loss for magnetic materials. To address this, we introduce an innovative approach that combines the Fast Fourier Transform (FFT) with a Feed-forward Neural Network (FNN), leading to a significant improvement in loss prediction accuracy. Our model, optimized through Multi-Objective Optimization (MOO), balances reduced complexity and heightened accuracy. To address data limitations, transfer learning is successfully employed for new materials with sparse data. This strategy offers the potential for improved training efficiency and broader applicability in the field.¹

I. INTRODUCTION

The empirical Steinmetz Equation (SE), introduced in 1890, has long served as the foundation for the standard modeling techniques for losses estimation in power magnetics. Despite various improvements such as iGSE and i2GSE, the accuracy of these curve-fitting techniques is still relatively low [1], [2]. The imprecise model will lead to a rough magnetic design. Thus, a large design margin is required to achieve an acceptable loss performance of the magnetic components.

One possible approach is using neural networks to accurately predict the loss of magnetic material, considering complicated material response mechanisms to different excitations and the factors involved, such as temperature and DC bias [3]. Different neural network structures have been proposed in [4], including FNN (Feed-forward Neural Network), which uses a four-layer FNN taking peak flux density B , duty ratio D , and fundamental frequency f to predict core loss density P_V . The disadvantage of this method is that the B waveform shape information cannot be included by only using the peak flux density value. One possible approach is to train different FNNs for different B waveforms. However, there are an unlimited number of B waveform shapes. Thus, it is impractical to categorize and then design separate FNNs for each case first.

¹ Z. Li (Z.Li-20@tudelft.nl), R. Mirzadaranji (R.Mirzadaranji@tudelft.nl), R. Liu (Ruijun.Liu@tudelft.nl), L. Wang (L.Wang-11@tudelft.nl), T. Luo (T.Luo-1@tudelft.nl), D. Lyu (D.Lyu@tudelft.nl), M. Ghaffarian Niasar (M.GhaffarianNiasar@tudelft.nl), and Z. Qin (Z.Qin-2@tudelft.nl) are with Department of Electrical Sustainable Energy, Delft University of Technology, Delft, 2628CD, The Netherlands.

A more effective method to represent the B waveform information, irrespective of the waveform's shape, is by applying FFT (Fast Fourier Transform). In this report, we utilize FFT to transform the waveform into harmonics. The magnitudes and frequencies of these harmonics, along with temperature data, are then input into the FNN for training. This approach limits the number of parameters and decreases the complexity; thus, the training time is considerably shorter in comparison to other more complex methods. To the best of the authors' knowledge, this is the first research to integrate FFT with FNN for estimating magnetic power loss.

To further improve the model performance, considering both accuracy and model size, a MOO (Multi-Objective Optimization) is carried out. To tackle the limited data size of the new materials, transfer learning has been taken to fine-tune the pre-trained model based on a large dataset of existing materials.

In this report, we detail a neural network modelling approach that integrates MOO and transfer learning to accurately forecast the power loss in magnetic materials under conditions of limited data availability. Accordingly, the report is organized as follows. Section II presents the employed FNN

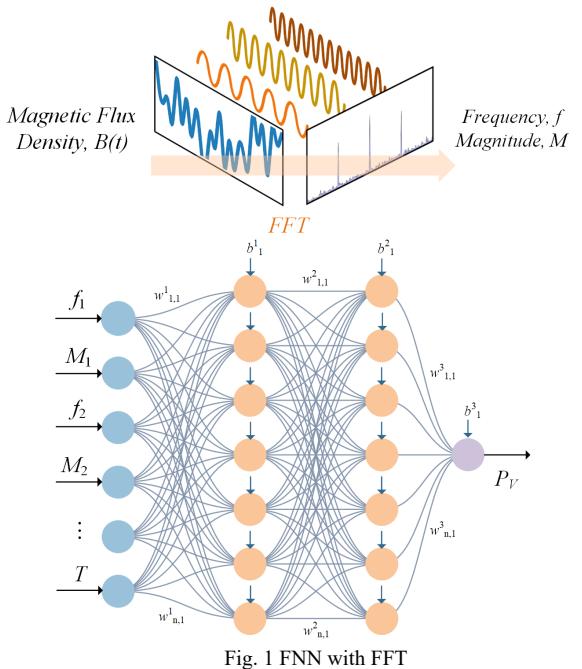


Fig. 1 FNN with FFT

architecture, elaborates on the rationale behind the chosen input format and the number of harmonics, and conducts a comparative analysis with other sophisticated AI methodologies. Section III details two approaches for predicting power loss in new materials: normal training and transfer learning, each integrating MOO. This section further compares the developed FNNs, aiming to identify the most effective model regarding both parameter count and accuracy. Finally, section IV concludes the report.

II. SELECTED NEURAL NETWORK AND COMPARISON WITH STATE OF THE ART

The FNN structure is shown in Fig. 1. As shown in the figure, the inputs of the FNN are $[f_1, M_1, f_2, M_2, \dots, f_n, M_n, T]$, where f_k and M_k represent k th harmonic frequency and magnitude, and this input structure is marked as case 0 for later comparison.

A. Selection of Input Format for FNN

Other input formats also have been considered, for example $[f_1, M_1, p_1, f_2, M_2, p_2, \dots, f_n, M_n, p_n, T]$, where p_k represents the phase information of the k th harmonic, and this case is marked as case 1; and $[M_1, M_2, \dots, M_n, f_1, T]$, where in this case only the fundamental harmonic frequency is included, and this is named as case 2. A comparison between these three cases has been carried out, as shown in Table 1.

Table 1
Compared Cases

Cases	Neurons in hidden layers	Total number of parameters	Avg. relative error on test set
0	(32,64,32)	4929	1.29
1	(64,64,64)	10433	2.2
2	(32,64,32)	4641	1.69

The test is carried out based on N27 material, and N27 data is split into 60%, 20%, and 20% for train, validation, and test datasets, respectively. Ten harmonics have been chosen for all three cases. The number of neurons in the hidden layers are determined based on the experience, and the activation function is ReLU. The training optimizer is set as Adam, and the model is trained in PyTorch. An exponentially decayed learning rate strategy is implemented to yield a better model convergence, where the initial learning rate is 0.004, and the decaying rate is 90% per 150 epochs [4]. In section III.A, a hyperparameter optimization will be carried out.

According to the comparison result, there are two conclusions. First, adding harmonics phase information does not help increase loss prediction accuracy. Second, when the model is trained using only a single fundamental frequency as input, its performance is inferior compared to when it is trained with both the frequencies and magnitudes of the harmonic components paired together. This observation holds under the scenario where the network parameters are initially selected based on empirical experience and not subject to further optimization. Therefore, the inputs for the FNN are chosen to be in the format of $[f_1, M_1, f_2, M_2, \dots, f_n, M_n, T]$.

An early stopping method is implemented in the training process to prevent overfitting and save computational resources. Early stopping halts training when a pre-set number of epochs pass without improvement in validation loss. It is used in the following model training process.

B. Selection of Number of Harmonics for FNN

A further question is how many numbers of harmonics are suitable to consider for model training and prediction. Three different FNN models are built for three, five, and ten harmonic inputs, respectively. A comparison between these three FNNs is shown in Table 2. The model training and testing settings are the same as the tests in Table 1.

Table 2
Comparison of Different Numbers of Harmonics as Input

No. harmonics	Neurons in hidden layers	Total number of parameters	Avg. error on test set	Max. error on test set
3	(16,32,16)	1217	1.38	7
5	(16,32,16)	1281	1.44	12
10	(32,64,32)	4929	1.29	17

The difference in neuron settings in the middle layer is due to the different number of inputs. As can be seen from the comparison results, though with the total number of parameters differences, the results of using three harmonics, five harmonics, and ten harmonics are similar. For the simplicity of the model structure, which is essential for training and future optimization, three harmonics are chosen for the input. Consequently, the input format for the FNN in this study has been established as $[f_1, M_1, f_2, M_2, f_3, M_3, T]$.

C. Comparison with Other Neural Network Structures

With the given magnetic flux density $B(t)$, magnetic field strength $H(t)$, frequency f , temperature T , and power loss P information, different AI models can be applied to predict the power loss. Apart from the scalar-to-scalar model presented in this report, [4] also offered two other models: sequence-to-scalar model using LSTM plus FNN and sequence-to-sequence model using encoder-decoder structure with LSTM. A detailed explanation of these two methods can be found in [4].

To compare with the models mentioned above, similar models, as shown in [4] are developed. The models are trained with N27 data, and for the LSTM+FNN, the input signal $B(t)$ is sampled at 1024-time steps per cycle, while for the encoder-decoder, it is down-sampled to 102-time steps per cycle to manage the higher computational demands. Each model is trained for 10000 epochs. As a result, a comparison between the proposed model, the LSTM+FNN, and the encoder-decoder structure is shown in Table 3. Additionally, the optimized FNN is detailed in section III.A is also shown for benchmarking. The neural network training was conducted in a PC equipped with an AMD Ryzen 7 5800H processor and an NVIDIA GeForce RTX 3060 graphics card.

The encoder-decoder is the most complex model, and the power loss estimation accuracy is unsatisfactory. The reason is that the power loss is not directly generated as the neural

network's output. On the contrary, the power loss is indirectly calculated by integration.

Table 3

Comparison between the FNN, LSTM plus FNN, and Encoder-Decoder

Cases	Total number of parameters	Training time	Avg. relative error on the test set
Initial FNN	1217	~1 hour	1.38
Optimised FNN	1419	~1 hour	0.75
LSTM+FNN	2809	~20 hours	1.37
Encoder-decoder	28097	~20 hours	3.34

The LSTM+FNN model achieves the highest accuracy without considering the optimization of the three models. Because it effectively captures detailed information from the B waveforms and directly generates power loss estimations. However, the sequence-to-scalar model presents two disadvantages compared to the scalar-to-scalar model. First, it is more complex and requires longer training time, which is unsuitable for further optimization. Second, in the case of transfer learning using the data measured by other research

groups, adjustments like down-sampling/up-sampling or modifying the model structure might be needed for transfer learning, especially when handling data sequences of different lengths.

In comparison, the proposed FNN model is the simplest and quickest for training. With further optimization, the model's performance can be increased significantly. Given its simplicity and high accuracy, the proposed FNN has been demonstrated to be suitable for the following development and optimization.

III. MULTI-OBJECTIVE OPTIMIZATION, TRANSFER LEARNING AND COMPARISON

The overall model construction and selection methods for the five new materials are presented in Fig. 2. In the normal training approach, for the five new materials, the FNN model for each material is first optimized with the MOO tool Optuna, as detailed in the next section. New materials data are split into 70%, 20%, and 10% for training, validation, and test purposes, respectively. Then, the best hyperparameter

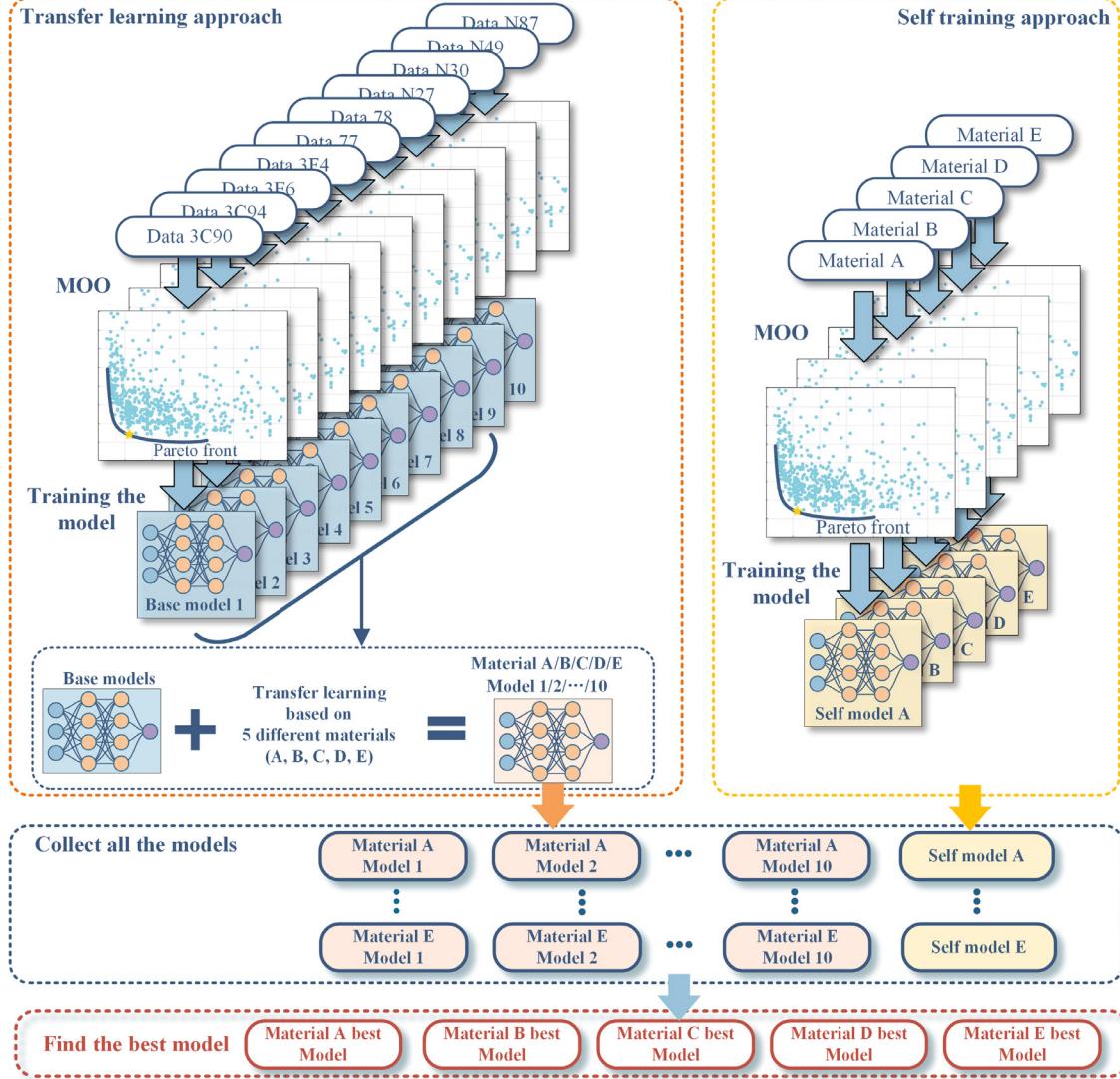


Fig. 2 Overall Model Training and Comparison Process

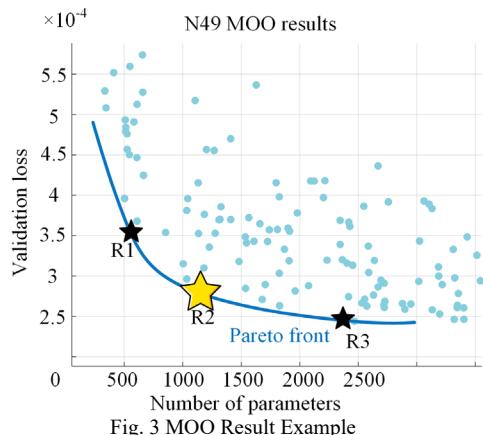
combination with a low number of parameters and low validation loss is selected. The test dataset performance is recorded for later comparison.

Although with MOO, the test performance is satisfactory with the normal training process, due to the low number of new material data, a better approach is by using transfer learning. This method holds significant importance in real-world scenarios, such as for designers who may lack adequate data on new materials required for normal training models. This part will be analyzed in III.B. The overall approach for transfer learning, as shown in Fig. 2, also starts with a MOO for the ten existing materials. This will increase the accuracy of the following step which uses transfer learning. Once the optimal model for each material is determined, the five new materials will be subjected to transfer learning across the ten models through fine-tuning. The ten existing materials data are split into 70%, 15%, and 15% for training, validation, and test purposes, respectively. This distribution strategy differs slightly from the previous one, as in the earlier data split, we aimed to maximize the use of the limited data available for training the new material. As a result, 50 new transfer-learned models for the five new materials will be acquired, and the test data results will be recorded. The transfer-learned and normal training results described previously are compared to select the final best model. We have also experimented with training the base model using a combined dataset of several materials (such as 3C90, N27, and 77). However, the outcomes of transfer learning with this approach did not surpass those achieved when training the base model exclusively with data from a single material (3C90).

A. Multi-Objective Optimization

In assessing the neural network model, both prediction accuracy and the number of parameters are crucial metrics. While accuracy reflects the model's effectiveness, the parameter count directly indicates its complexity and significantly impacts the time required for training and fine-tuning. To optimize the model considering both values, a MOO is needed.

MOO aims to find the optimal set of parameters, simultaneously optimizing multiple objectives, leading to trade-offs and compromises. Optuna, a popular Python library,



provides a comprehensive framework for conducting MOO. Utilizing Optuna for MOO, we focus on two primary objectives: reducing the validation loss to enhance model accuracy and minimizing the number of parameters. NSGA-II (Non-dominated Sorting Genetic Algorithm II) is favored as the optimization engine for its effective balance in ranking and diversity preservation in MOO contexts. The search space is defined as follows: the number of middle layers in the FNN structure varies between 2 and 5, each containing 8 to 64 neurons. The activation functions considered include ReLU, LeakyReLU, Tanh, Sigmoid, ELU, SELU, and SiLU. Possible optimizers are Adam, SGD, RMSprop, and AdamW. The batch size options are set at 64, 128, or 256.

For each of the five new materials, MOO involves 500 trials, each comprising 3000 epochs. In contrast, for the ten existing materials, which have larger datasets, MOO consists of 200 trials with 1000 epochs each. An example MOO result is presented in Fig. 3, where stars indicate the best hyperparameter combinations. Table 4 summarizes these combinations, and the optimizer for three cases is Adam. The N49 test dataset demonstrates that accuracy increases with the number of parameters. However, to strike a balance between the number of parameters and accuracy, R2 is selected as the optimal choice for further transfer learning due to its proximity to the knee point of the Pareto front. A similar analysis approach is applied to the other materials.

Table 4
Three Different Points on N49 MOO Pareto Front

case	Batch size	Hidden layers neurons	No. params.	Act. Func.	N49 avg.	Material E TF avg.
R1	64	(21, 17)	560	Tanh	2.01	2.75
R2	128	(29, 29)	1132	Tanh	1.9	2.49
R3	128	(38, 19, 33, 19)	2371	SiLU	1.46	2.35

B. Transfer Learning

After the MOO, ten base models trained on existing materials are ready for transfer learning. New material data will be directly input into the pre-trained base model. This process will fine-tune the neural network's parameters, adapting the model to accommodate the characteristics of the new materials.

A test is carried out to testify the MOO results and the transfer learning effectiveness. Based on Table 4 N49 trained models, the new material E data are applied for fine tuning. The transfer learning results are shown in the last column of the table. It can be similarly concluded that an increase in the number of parameters enhances the accuracy of transfer learning results. To strike an optimal balance between the number of parameters and accuracy, the knee point on the Pareto front is identified as the ideal combination of hyperparameters.

C. Normal Training Results and Transfer Learning Results Comparison

Based on the overall process shown in Fig. 2, the results of the transfer learning approach and the normal training

Table 5
Ten Existing Materials Optimized MOO Models Transfer Learning Results

material (number of parameters)	Self-training result		Transfer learning results									
	Average error	95% error	Material A		Material B		Material C		Material D		Material E	
			Average error	95% error	Average error	95% error	Average error	95% error	Average error	95% error	Average error	95% error
N27(1419)	0.75	1.97	2	6.7	0.75	1.89	1.39	3.93	3.16	10	2.4	7.25
N30(2197)	0.41	1.1	2.08	7.85	0.7	1.65	1.1	3.24	3.7	10	2.28	7
N49(1132)	1.9	5.78	2.688	8.9	0.79	2.2	1.37	3.9	5.11	19	2.49	7
3F4(2117)	1.45	3.8	2.58	7.8	0.84	2.17	1.23	3.4	5.2	15.7	2.66	10.9
3E6(1295)	0.455	1.19	2.25	6.3	0.74	1.98	1.23	3.62	4.26	10.65	2.77	8
3C90(2705)	0.83	2.35	2.23	8.1	0.76	1.97	1.2	3.78	3.86	10.9	2.38	9.24
3C94(3021)	0.75	2	3.1	10.4	0.75	2.02	1.37	4.2	5.9	22.4	2.63	7.09
N87(1525)	0.74	2	2.54	7.67	0.73	1.9	1.19	4	4.6	14.9	2.74	10.7
77 (2454)	0.77	2.1	2.2	6.37	0.92	2.48	1.21	3.54	4.77	14	2.2	6.4
78(2285)	0.72	1.89	2.19	7.1	0.69	1.9	1.21	3.72	5.23	15.38	2.69	8.17

Table 6
Comparison of the Best Transfer Learned Results and Regular Trained Results for the Five New Materials

New Material	Self-training result			Transfer learning result			
	Number of parameters	Average error	95% error	Number of parameters	Average error	95% error	Base model
A	1662	2.18	6.73	1419	2	6.7	N27
B	552	0.9	2.2	2197	0.7	1.65	N30
C	935	1.65	4.4	2197	1.1	3.24	N30
D	1857	6.3	21.8	1419	3.16	10	N27
E	1551	2.47	7	2454	2.2	6.4	77

approach for the five new materials are presented in Table 5 and Table 6. Table 6 compares the transfer learning results for the five new materials, and the best results are highlighted in green. Table 7 presents test dataset accuracy for the MOO models for the five new materials to compare with the transfer learning results.

As seen in Table 6, from an accuracy perspective, the results of transfer learning significantly outperform those of self-training, mainly due to the limited dataset available for the new materials. Although the number of parameters in transfer learning approaches is slightly higher than in self-training methods, this is not a significant concern for applying transfer learning to new materials. The rapidity of the transfer learning process, as demonstrated in our tests, enables quick adaptation to the characteristics of new materials. Consequently, despite the larger parameter size, the speed of transfer learning makes it a more practical choice. Therefore, considering both the superior accuracy and the efficiency in learning new material properties, the transfer learning models are deemed the most effective for new materials.

The number of parameters for the final model for the five new materials is summarized in Table 7 for the committee's convenience.

Table 7
Number of Parameters for New Materials

Material	A	B	C	D	E
Number of parameters	1419	2197	2197	1419	2454

IV. CONCLUSION

This report comprehensively presents a neural network based modelling approach for accurately predicting the power

loss of magnetic materials with limited data, incorporating MOO and transfer learning techniques. The fundamental neural network employed in this study is a FNN. Through a detailed comparison of various input structures and the number of harmonics, the chosen inputs are the magnitudes and frequencies of the first, second, and third harmonics, plus temperature. Compared to other advanced methods, such as LSTM or encoder-decoder models, the proposed method achieves satisfactory accuracy with lower complexity. The model's performance is further enhanced by MOO, targeting both a reduction in the number of parameters and an increase in accuracy. Finally, a comparative analysis between the transfer learning and self-training approaches, both integrated with MOO, is conducted for new materials with limited data. The transfer learning shows a better performance with considerable lower error and marginal increase in the number of parameters.

REFERENCES

- [1] F. Kral, "Analysis and Implementation of Algorithms for Calculation of Iron Losses for Fractional Horsepower Electric Motors," Master's Graz University of Technology, 2017.
- [2] C. N. M. Smailes, R. Fox, J. Shek, M. Abusara, G. Theotokatos, P. McKeever, "Evaluation of core loss calculation methods for highly nonsinusoidal inputs," presented at the IET International Conference on AC and DC Power Transmission, 2015.
- [3] D. Serrano. e. al., "Why MagNet: Quantifying the Complexity of Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14292-14316, 2023, doi: 10.1109/TPEL.2023.3291084.
- [4] H. Li. e. al., "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15829-15853, 2023, doi: 10.1109/TPEL.2023.3309232.

Advancing Power Magnetic Materials Modeling through Attention-based U-Net with Linear Conditioning

Syed Irfan Ali Meerza, Kody Froehle, Han (Helen) Cui, Daniel Costinett, Jian Liu .

Abstract—The evolution of power electronics is fundamentally constrained by the performance of magnetic components, accounting for substantial cost and loss in power converters. The MagNet Challenge spearheads the quest for breakthroughs in the modeling and design of power magnetics, where traditional methods like the Steinmetz Equation, despite upgrades, fall short in accuracy and adaptability. This report proposes a novel and data-driven framework that synergizes empirical modeling with contemporary computational techniques. Our approach seeks to refine the predictive precision of core loss in magnetic materials under varying operational conditions and to expedite the design process, significantly reducing development cycles and reliance on extensive engineering expertise. Central to our methodology is a sophisticated deep learning model, innovatively based on a metadata-conditioned U-Net architecture enriched with attention mechanisms. This model is specifically designed to adapt to the intricate and varying nature of magnetic materials and operational environments. We further refine the model’s robustness through noisy-supervision-based self-training, a strategy that ensures its effectiveness on unseen test data. By transcending the constraints of traditional magnetic component modeling, the proposed framework of power magnetic materials modeling not only offers a more efficient and tailored design process but also sets the stage for the next generation of power electronics.

I. INTRODUCTION

Magnetic components are lossy and bulky in power converter systems, which has become a main bottleneck for the new generation of high-performance power conversion systems. The non-linear nature of magnetic materials and the variability introduced by manufacturing processes thwart traditional modeling efforts. Current empirical models, while foundational, offer limited accuracy, and the dynamic operational conditions—spanning temperature, frequency, and DC bias—exacerbate the challenge. Consequently, the power electronics field yearns for a rapid and precise modeling technique that can streamline the design process and reduce the reliance on extensive engineering expertise.

In addressing the MagNet Challenge, our team has developed a hybrid machine learning framework, integrating several advanced components, including data augmentation via Generative Adversarial Networks (GAN) [1], metadata-conditioned attention-based U-Net [2], noisy supervision [3],

Syed Irfan Ali Meerza, Kody Froehle, Daniel Costinett and Jian Liu are with Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville (UTK), TN, USA. Emails: smeerza@vols.utk.edu, kfroehl@vols.utk.edu, dcostine@utk.edu, jliu@utk.edu.

Han (Helen) Cui is with Tianjin University, China. She is also an Adjunct Assistant Professor at UTK. Email: helencui@utk.edu

and knowledge distillation [4]. The framework initiates with the implementation of GAN-based data augmentation, tailored to fulfill the enhanced data volume requisites imperative for efficient model training. Subsequently, the framework augments the attention-based U-Net [5] by assimilating metadata (e.g., temperature, frequency, and waveform type) through the application of cost-effective affine transformations on feature maps.

Moreover, a pivotal element of our methodology is the incorporation of a self-supervised learning paradigm. This paradigm is primarily focused on exploiting the latent supervisory signals inherent in unlabeled data, particularly from testing materials, to refine the model’s adaptability to new datasets. Furthermore, we employ knowledge distillation techniques to effectively reduce the model size, ensuring a more efficient and deployable solution. The overarching objective is to realize substantial performance gains with minimal manual intervention, leveraging the intrinsic supervisory characteristics within the testing materials for model refinement.

II. PRELIMINARIES

A. 1-Dimensional Generative Adversarial Network (GAN)

Generative Adversarial Networks (GANs) in a one-dimensional (1-D) setting are a simplified version of the standard GAN framework. They consist of two neural networks, the Generator (G) and the Discriminator (D), that are trained simultaneously through adversarial processes. The generator aims to produce data that is indistinguishable from real data, while the discriminator strives to accurately distinguish between real and generated data. The adversarial training of these networks is orchestrated by a loss function that encapsulates their tug-of-war:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x_{p_d}} [\log D(x)] + \mathbb{E}_{z_{p_z}} [\log(1 - D(G(z)))] \quad (1)$$

where $\mathbb{E}_{x_{p_d}} [\log D(x)]$ represents the expected log-probability that the discriminator assigns to real data being authentic, while $\mathbb{E}_{z_{p_z}} [\log(1 - D(G(z)))]$ captures the expected log-probability that the discriminator assigns to fake data being inauthentic. The generator’s goal is to minimize this function, thereby generating data that is increasingly likely to be classified as real by the discriminator. In our approach, we leverage 1D GAN to expand limited datasets, allowing for the development of models that predict magnetic behavior with enhanced accuracy.

B. U-Net with GRU

Integrating Gated Recurrent Units (GRUs) [6] into a U-Net architecture for sequence-to-sequence (*seq2seq*) generation represents a significant adaptation from its traditional applications. In this variant, both the encoder and decoder components of the U-Net are constructed using GRU layers. This design allows the model to effectively handle sequential data, such as Magnetic Flux Density and Magnetic Field Intensity.

In the encoder part of the U-Net, GRUs process the input sequence, creating a compressed representation that captures the temporal context and dependencies within the data. This encoded information is then passed through the bottleneck of the network and subsequently fed into the decoder. The decoder, also composed of GRU layers, aims to reconstruct or transform the sequence into the desired output format, maintaining the sequence's integrity and contextual relevance. The skip connections, a hallmark of U-Net architecture, are adapted here to connect corresponding GRU layers in the encoder and decoder, allowing for the transfer of temporal context directly across the network, which is particularly beneficial for complex *seq2seq* tasks.

C. Knowledge Distillation

Knowledge distillation [4] is a technique in machine learning where a smaller, less complex model (the “student”) is trained to replicate the behavior of a larger, more complex model (the “teacher”). This process involves using the outputs of the teacher model as an additional training signal for the student. The rationale behind knowledge distillation is that the student model, while smaller and more computationally efficient, can learn to perform similarly to the teacher by mimicking its output patterns. This approach is particularly useful for deploying models in resource-constrained environments like mobile or embedded systems, where the balance between model size and performance is critical.

The process typically involves training the teacher model to its full capacity on the training data and then using its predictions to train the student model. The student learns to mimic the teacher’s outputs by minimizing a distillation loss function that signifies the divergence between the teacher’s and student’s output distributions. The student’s loss function is like below:

$$\min \mathcal{L} = \alpha \cdot H(y, z_S/T) + (1 - \alpha) \cdot D_{KL}(z_T/T, z_S/T), \quad (2)$$

where \mathcal{L} represents the total loss, H denotes the Mean Squared Error (MSE) loss, y is the true output, z_S and z_T are the outputs from the student and teacher models respectively, T is a temperature parameter utilized for softening the probabilities, D_{KL} signifies the Kullback–Leibler divergence, and α is a balancing factor for the two components of the loss function.

III. FRAMEWORK OVERVIEW

As shown in Fig. 1, our framework involves a four-stage process, each stage building upon the insights and outputs of the previous one. This structured approach ensures a comprehensive development and refinement of the model, tailored to

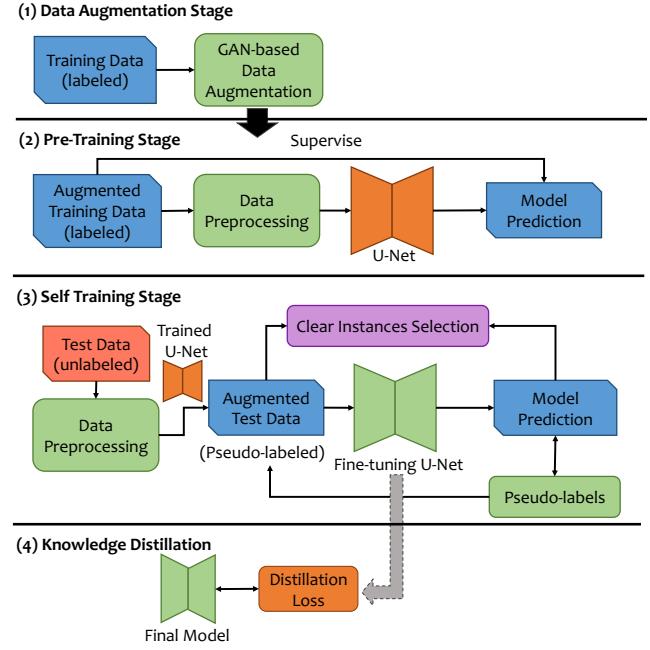


Fig. 1. Framework Overview.

effectively handle diverse and complex datasets. The stages are as follows:

(1) Data augmentation stage. Initially, we focus on diversifying the training data through the data augmentation stage. We utilize a 1D GAN, trained on the provided training dataset, to create additional data points. These synthetic points closely resemble the original data distribution, thus expanding the dataset and boosting the model’s capability to generalize to new, unseen data.

(2) Pre-training stage. Following augmentation, the augmented dataset is then utilized to *pre-train* an Attention-based Affine Transformation (AFT) U-Net (Section IV-B). This stage is standard supervised learning, where the model is trained using labeled, augmented training data.

(3) Self-training stage. Due to the potential distribution bias between the training and test datasets, the model trained on the training set might not provide optimal performance on the test set. This stage is used for elevating the pre-trained model to higher performance levels on the test dataset. Specifically, in this stage, we will utilize the pre-trained model’s predictions on the test data as pseudo-labels. This approach is designed to fine-tune the model, thereby tailoring it more effectively to the characteristics of the test set. A small-loss strategy is employed to select the most reliable instances and their corresponding pseudo-labels, which then guide the fine-tuning process. The iterative refinement of the model parameters and pseudo-labels during this stage results in a significantly enhanced model, specifically optimized for the test dataset.

(4) Knowledge distillation stage. The final stage involves knowledge distillation, where the insights and capabilities of the fine-tuned, larger model are transferred to a smaller, more efficient variant. In this process, the larger model serves as a ‘teacher’, providing outputs for both the training and test datasets, while the smaller ‘student’ model learns to replicate

these outputs. This knowledge transfer is crucial for creating a compact model that retains the high performance of its larger predecessor.

Together, these stages represent a holistic and effective strategy for model development, ensuring that each phase contributes to the creation of a robust, high-performing model adept at handling the complexities of material data.

IV. FRAMEWORK DESIGN

A. GAN-based Data Augmentation

Limited training data often poses a significant challenge in machine learning, particularly for models like U-Net, which can become prone to overfitting where the model learns the specifics of the training data, including its noise and outliers, too precisely, leading to poor generalization on new, unseen data. This issue is especially acute when the available training dataset is small, as is often the case. To address this, we employed a 1D Generative Adversarial Network (GAN) to generate synthetic data samples, enhancing our dataset's diversity and size. This approach ensures that the synthetic data closely mirrors the real data's distribution.

Given that our dataset comprises three types of waveforms, we opted for a conditional 1D GAN, allowing us to exert control over the generated data based on the type of waveform. This conditionality is crucial for maintaining consistency and relevance in the synthetic data. During the GAN's training phase, we meticulously concatenated various components – the B signal, H signal, frequency, and temperature – into a single vector. This comprehensive approach ensures that the generated data not only mimics the real data in terms of distribution but also respects the intrinsic properties and relationships inherent in the original dataset. As a result, the augmented dataset provides a more robust foundation for training the U-Net model, significantly improving its performance and generalizability on unseen data.

B. Attention-based U-Net with Linear Conditioning

In enhancing the sequence-to-sequence (*seq2seq*) capabilities of the U-Net architecture, we leverage the Affine Transformation technique, an innovative approach originally used in visual question answering. This technique significantly augments the network's ability to integrate and process metadata, a crucial element in our application, which includes information such as frequency, temperature, and the type of input wave (sine, triangular, or trapezoidal). To accurately determine the type of input wave, we take the second-order derivative of the input signal. This technique yields distinct responses for different waveforms—a sine wave response for sinusoidal inputs, two spikes for triangular waves, and more than two spikes for trapezoidal waves. By counting these spikes, we can effectively identify the type of the input waveform. This metadata integration is key to achieving more accurate and contextually relevant *seq2seq* generation.

Our model's architecture is based on a 2D U-Net as shown in Fig. 2, designed to handle dual inputs: magnetic flux density and the one-hot encoded metadata. The Affine Transformation (Aft) layers, in synergy with a dedicated generator, are tasked

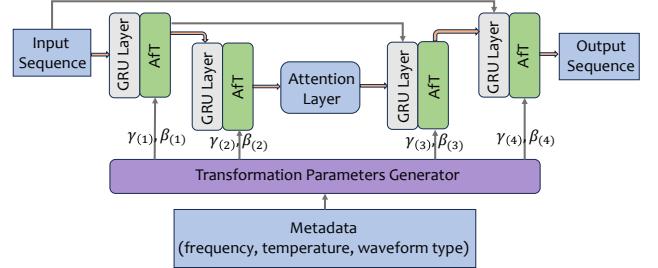


Fig. 2. Attention-based U-Net with Linear Conditioning.

with conditioning the neural network using this metadata. Central to the Aft layers' functionality are the $\gamma(i)$ and $\beta(i)$ parameters, which linearly modulate the inputs of each Aft layer. These parameters are generated by a specialized Transformation Parameters Generator, featuring two hidden layers with 64 and 16 neurons, respectively. The generated γ and β values, unique for each filter, are applied to modulate the GRU feature maps within the Aft layers, ensuring that each layer's output is finely tuned in accordance with the metadata. Shared weights across the generator enhance learning efficiency, maintaining a consistent feature extraction from the metadata. Additionally, the sigmoid activation confines the values between 0 and 1, ensuring a controlled modulation effect.

In addition, an attention layer has been strategically integrated to bridge the encoder and decoder part of the U-Net. This layer refines the model's focus, allowing it to focus on salient features and sequences. It enhances the network's performance in handling complex dependencies and contextual variations, which is particularly beneficial in our seq2seq tasks.

C. Self-training Adaptation

To improve the model's ability to generalize to new, unseen test data, we employ the concept of noisy supervision [3]. This semi-supervised learning method is designed to leverage unlabeled test datasets in this MagNet challenge, enabling the model to better adapt to these test sets. This approach is highly beneficial in scenarios where there is a mix of labeled and unlabeled data. In our case, the training data are labeled while the testing data remain unlabeled. Noisy supervision facilitates learning from both labeled and unlabeled data, enhancing the model's adaptability. This is crucial in addressing the challenge of domain shift, a prevalent issue where a model trained in one setting underperforms in a different one.

In this process, the pre-trained model, trained on the labeled training data, attempts to generalize its learning to the target domain (i.e., test sets) through self-training. Particularly, the model first performs regular model prediction on the test datasets and generates pseudo-labels (i.e., H signals). Initially, the model's performance on the test dataset is suboptimal due to the potential distribution bias between the training and test datasets. We thus purify the obtained pseudo-labels through a clean instances selection strategy, choosing only the fraction of the instances that produce small losses to supervise the self-training process.

In addition, deep networks tend to initially fit into clean or “easy” instances in the early training epochs and later start adapting to noisy or “hard” instances. To optimize this training process, we introduce $R(T)$, which adjusts the fraction of data instances utilized for self-training at each training epoch T . Initially, $R(T)$ is set to 1, meaning all instances are included, as early in training, deep networks are less prone to memorize noise, making the full dataset beneficial for learning. However, as training advances, $R(T)$ is systematically decreased. This is in response to the tendency of networks to begin overfitting to noisy data, typically indicated by higher losses. By progressively reducing $R(T)$, we prioritize training on cleaner, more reliable data, effectively mitigating the risk of the networks overfitting to noisy labels with increasing epochs.

Our strategy for selecting clean instances in the self-training stage is crucial in reducing error rates in pseudo-labels, ensuring cleaner supervision. This ongoing refinement of pseudo-labels, coupled with repeated learning from the test dataset, enables the model to progressively achieve satisfactory performance, adeptly adapting to the specific features of the target domain.

D. Framework Workflow

Our methodology, meticulously tailored for each type of material, employs a four-stage training process to optimize model performance for each specific material. The initial stage, focused on augmentation, leverages a 1D GAN network to enhance the raw training data. This involves training the GAN to replicate the unique distribution of the training data, followed by generating 3,000 to 5,000 synthetic data samples specific to each material type. These samples are then amalgamated with the original training data, creating a comprehensive, augmented training dataset. This enriched dataset serves as the foundation for pre-training an Attention-equipped Affine Transformation (Aft) U-Net model, setting the stage for subsequent steps.

The second stage continues with pre-training the U-Net model using the augmented dataset. Before training, we normalize the dataset using the mean and variance of the dataset. This training is a standard supervised learning where the model is trained using labeled augmented data.

In the third stage, we infer the pre-trained U-Net model on the test dataset to extract pseudo-H signals. These signals, once integrated with the test dataset, form an augmented test dataset. This dataset, combined with the augmented training dataset, culminates in a custom dataset tailored for each material, bridging the gap between training and real-world application. This stage involves fine-tuning the pre-trained U-Net model through a self-training approach. Here, the custom dataset is used to refine the model, focusing on updating it based on select low-loss instances. This selection is strategically governed by an instance selection algorithm, which dynamically adjusts the number of instances chosen, starting with a higher count and progressively reducing it.

In the final stage, we implement knowledge distillation to impart insights from the fine-tuned U-Net model to a smaller, more efficient version of the Aft U-Net with atten-

TABLE I
NUMBER OF MODEL PARAMETERS

Material	A	B	C	D	E
Number of Parameters	23,000	23,000	23,896	32,546	25,990

tion. This ensures that the smaller model maintains the high-performance characteristics of the larger model while being better suited for scenarios demanding computational efficiency. This comprehensive, multi-stage approach not only facilitates effective learning from the available data but also optimizes the model for enhanced performance and efficiency, particularly on unseen data.

V. MODEL SIZE

Table I shows the number of parameters for the five models we developed, each corresponding to one of the five materials.

VI. CONCLUSION

In conclusion, our hybrid machine learning framework represents a significant stride forward in addressing the challenges posed by the non-linear nature and variability of magnetic materials in power conversion systems. By judiciously combining advanced techniques like GAN-based data augmentation, metadata-conditioned attention-based U-Net, noisy supervision, and knowledge distillation, we have crafted a solution that not only enhances modeling accuracy but also simplifies the design process in power electronics. The incorporation of self-supervised learning paradigms and the exploitation of latent supervisory signals in unlabeled data are particularly notable for their roles in improving the adaptability of the model to new datasets. This framework adeptly handles the dynamic operational conditions and variability inherent in magnetic components, reducing the reliance on extensive empirical models.

REFERENCES

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [2] A. Lemay, C. Gros, O. Vincent, Y. Liu, J. P. Cohen, and J. Cohen-Adad, “Benefits of linear conditioning with metadata for image segmentation,” *arXiv preprint arXiv:2102.09582*, 2021.
- [3] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, “Co-teaching: Robust training of deep neural networks with extremely noisy labels,” *Advances in neural information processing systems*, vol. 31, 2018.
- [4] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, 2015.
- [5] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pp. 234–241, Springer, 2015.
- [6] R. Dey and F. M. Salem, “Gate-variants of gated recurrent unit (gru) neural networks,” in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pp. 1597–1600, IEEE, 2017.

MagNet Challenge 2023-Ferrite Core Loss Model Based on LSTM and Transfer Learning

Zhanlei Liu, Cao Zhan, Yongliang Dang, Yukun Zhang, Na Wang, Yiting Chen, Yiming Zhang

Abstract- The accurate modelling of core loss is significant for the temperature prediction and thermal design of magnetic components. Existing core loss calculation methods for ferrite are either inaccurate or time-consuming, hindering the precise design of magnetic components. In this paper, a neural network based ferrite core loss model is proposed. The CNN+LSTM network is proposed to model the core loss on large datasets. To reduce the demand for data size, a transfer learning network is utilized to model the core loss on small dataset. The proposed LSTM and transfer learning based network can achieve accurate core loss calculation.

I. INTRODUCTION

High frequency transformers (HFT) and high frequency filter inductors (HFFI) are widely used in DC-DC converters such as grid-connected inverters, railway transports and power supply of data centers. Magnetic components are key components in HFTs and HFFIs, which provide the roles of power transmission and energy storage. The modelling and design of magnetic components is essential in the design of HFTs and HFFIs.

The core loss can account for half of the total power loss in magnetic components. Accurate modelling of core loss is one of the key steps in the modelling of magnetic components. Various core loss calculation methods had been proposed. The Original Steinmetz equation (OSE) proposed by Steinmetz is simple and practical, which is limited to sinusoidal excitations. In order to extend the empirical model to high frequency and non-sinusoidal excitations, some modified Steinmetz equations were proposed, including the MSE, IGSE, WcSE, I²GSE, and CWH. However, the accuracy of Steinmetz equations in wide ranges of frequency and flux density is low. The loss separation method proposed by Bertotti cannot achieve accurate core loss calculation for ferrite under high frequency excitations. The hysteresis models are too complex to apply in practical engineering. The neural network is a data-driven model, which not relies on the specific core loss mechanisms. The researchers in Princeton and Dartmouth have collected large ferrite core loss data and established some neural network models to predict the ferrite core loss such as feedforward neural network, transfer learning and long short-term memory network models. The models can predict the core loss under arbitrary waveform excitations and the predicted core loss showed higher accuracy than traditional empirical equations. However, the model complexities are high and the core loss mechanisms are not explained in the models.

In this paper, a neural network based ferrite core loss model is proposed. The CNN+LSTM network is proposed to model the core loss under large core loss datasets. The transfer

learning method is utilized to model the core loss under small core loss datasets. The accuracy of these neural network models are analyzed.

II. CNN+LSTM NETWORK BASED CORE LOSS MODEL ON LARGE DATASET

A. Analysis of the Large Core Loss Dataset

The core loss dataset used in this model is established by the researchers from Dartmouth and Princeton Universities, including various materials (3C90, 3C94, 3E6, and et al.), wide frequency (50kHz-500kHz), wide flux density (10mT-300mT), wide temperature (25°C-90°C) and several excitation waveforms (sinusoidal, triangle, and trapezoidal). The datasets include flux density sequences, frequencies, temperatures and core loss densities.

B. CNN+LSTM Network

A CNN+LSTM network is proposed to model the core loss as shown in Fig. 1. The model takes the flux density sequence, temperature, and frequency as input and the core loss density as output. The CNN is used to extract the local features in the flux density sequence and the LSTM is used to extract the time series features in the flux density sequence. The FCs are used to characterize the core loss density with flux density sequence features, frequency and temperature.

In this model, the input layer has a feature dimension of 16. The CNN module consists of a convolutional layer, ReLU activation function, and max pooling layer, using 1D convolution for feature extraction. It has 1 input channel, 16 output channels, a convolution kernel size of 25, and a pooling window size of 10. This module is used to extract and integrate features from multiple batches of time series data. The LSTM layer models the input time series data to extract higher-level feature representations. The use of Bi-LSTM helps capture past and future contexts of input elements, maintaining long-term dependency relationships between model parameters and learning data. Additionally, it helps handle variable-length sequences and can process sequences of different lengths in batches. The input size of the feature dimension after convolutional kernel pooling is 16. The hidden size determines the output dimension of the LSTM and increases model complexity and training time. A larger hidden size can provide richer feature expression capability. The hidden size of this model is 32. The number of layers determines the depth of the LSTM network, allowing for more complex sequence patterns. This model has 1 layer. The final output size of the Bi-LSTM is 64. After the first fully connected layer with the activation function ReLU, the output

size is 32. The magnetic field intensity features extracted from the previous steps are then combined with peak intensity, frequency, and temperature to form a sequence of size 36. After the second fully connected layer with the activation function ReLU, the output size is 16. Finally, the loss value is obtained through the last fully connected layer. Dropout controls the regularization level of the model, and a larger dropout can reduce overfitting to some extent and reduce the training burden of the model. In addition, the batch size is 32, the number of epochs is 200, and the learning rate is 0.01. The CNN has a total of 5068 parameters, including the convolutional and pooling layers, while the Bi-LSTM has a total of 54784 parameters, including the hidden and fully connected layers.

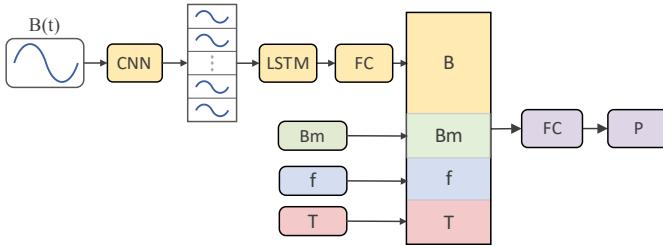


Fig. 1 Histograms of core loss error under large datasets. (a) 3C90, (b) 3E6, (c) 77, (d) N27.

C. Results

The errors of the proposed core loss model are shown in Fig. 2. The mean errors are within 3%, while error of the top 95% of the data are within 5%. Particularly for 3E6, the mean error is about 0.87%, while the error of the top 95% of the data is below 2.5%.

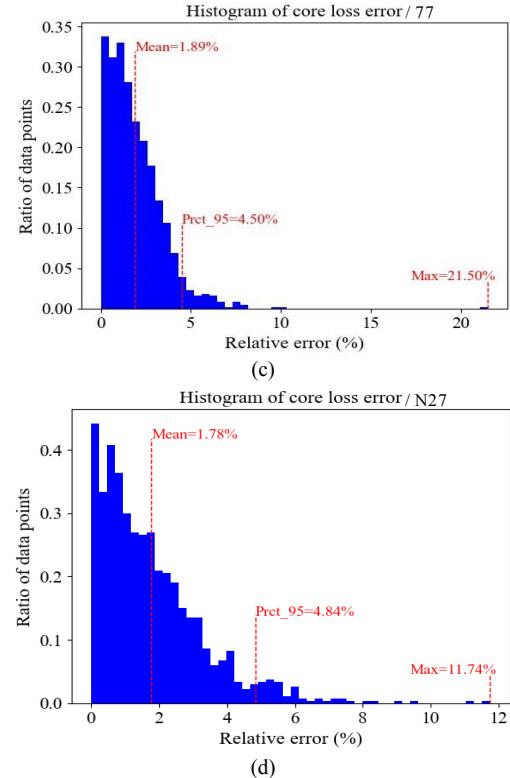
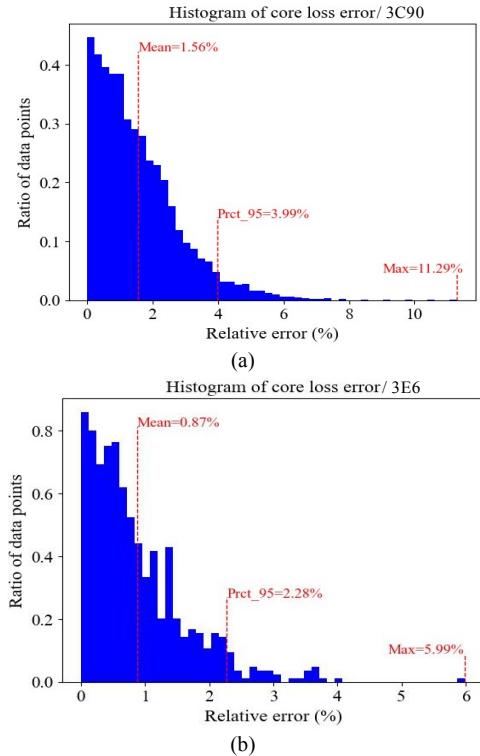


Fig. 2 Histograms of core loss error based on large datasets. (a) 3C90, (b) 3E6, (c) 77, (d) N27.

III. TRANSFER LEARNING BASED CORE LOSS MODEL ON SMALL DATASET

The neural network model based on CNN+LSTM can realize accurate modeling and calculation of ferrite core loss. However, the above model has high requirements on the number of samples in the training set and needs to be modeled based on loss data stimulated by wide frequency, wide magnetic flux density, wide temperature and various complex waveforms, which is difficult to be applied in practical engineering. Therefore, how to use small sample data set to achieve core loss modeling is of great significance for practical engineering.

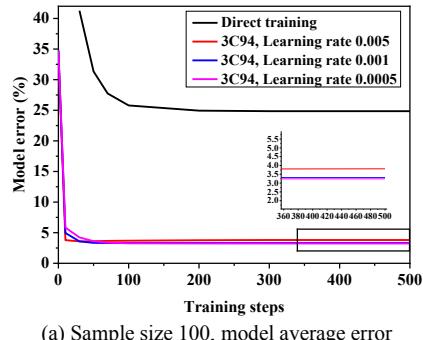
Transfer learning (TL) is a new machine learning method developed in recent years. It uses the source task to pre-train the model, and then fine-tunes the pre-trained model through the target task to obtain the learning model of the target task. It can be used to solve the problem of insufficient data. Compared with traditional machine learning, it is more efficient and accurate. The TL model based on parameter fine-tuning is simple and fast in calculation, and is a common method in TL. The fine-tuning method first trains a pre-trained model on a large sample data set, then freezes the network parameters of the first few layers of the pre-trained model, and trains and fine-tunes the basic network parameters behind the pre-trained model based on the new small sample data. Finally, the pre-trained model is migrated to new data samples.

TABLE I
THE PRE-TRAINED MODEL IS DIRECTLY APPLIED TO THE PREDICTION ACCURACY OF 3C90 IN A LARGE SAMPLE DATASET

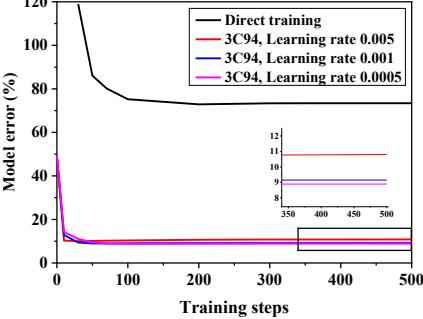
Pretraining model	3C94	3E6	3F4	77	78	N27	N30	N49	N87
Model mean error /%	11.1	165.8	113.74	19.79	15.05	32.44	137.36	52.89	48.8
Model 95% error /%	28.54	461.01	324.3	35.43	29.52	52.07	331.11	143.41	67.03

In order to make the small sample dataset of five new materials achieve better prediction effect, this paper established a ferrite loss model based on TL on the basis of the large sample data set neural network model and the similar mechanism between different types of ferrite core losses. In order to verify the validity of the parametric fine-tuning loss prediction model, this paper takes the 3C90 model ferrite as an example to construct 5 small sample datasets, the sample sizes of which are 100, 200, 400, 600 and 1000 respectively. The test set samples are the remaining samples after the total data set samples are removed from the training set samples. Then evaluate the prediction accuracy of nine other models of ferrite core loss pre-training models directly applied to 3C90, as shown in TABLE I below. It can be seen that 3C94 ferrite core loss pre-training model has the highest accuracy when applied to 3C90. Therefore, the training results of 3C94 ferrite core loss model will be used as the pre-training model below, and 3C90 small sample data set will be used to fine-tune the parameters of the pre-training model to achieve model migration.

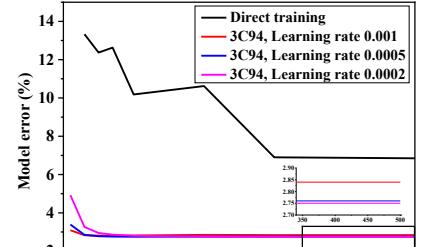
For the 3C90 ferrite core loss small sample data set, the model errors of direct training and TL training were compared, as shown in Fig. 3. It can be seen that for small sample data sets, TL can greatly improve the convergence speed and accuracy of the model, and better model performance can be obtained by selecting an appropriate learning rate. In addition, with the increase of the number of training samples, the accuracy of the model is improved.



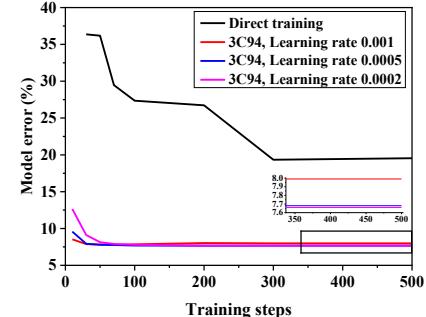
(a) Sample size 100, model average error



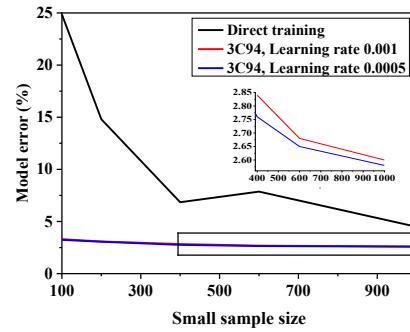
(b) Sample size 100, 95% error of the model



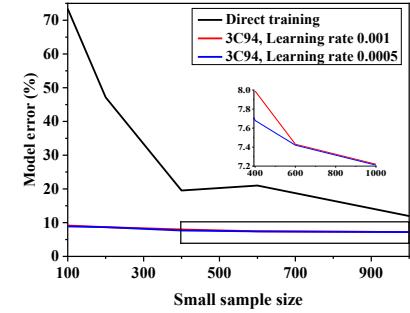
(c) Sample size 400, model average error



(d) Sample size 400, 95% error of the model



(e) Model average error



(f) 95% error of the model

Fig. 3 Comparison of model errors between direct training and TL training in small sample dataset.

Therefore, in order to make the loss prediction of the 5 unknown new materials more accurate, the prediction

accuracy of other 10 known models of ferrite core loss pre-training model directly applied to the 5 new materials was evaluated first, and the training results of the known model of ferrite loss with the highest prediction accuracy were selected as the pre-training model of the unknown new materials. Finally, the sample data set of the unknown new material was used to fine-tune the parameters of the pre-trained model to achieve model migration and accurate prediction of the loss of the new material with small sample data set. The specific process is shown in Fig. 4.

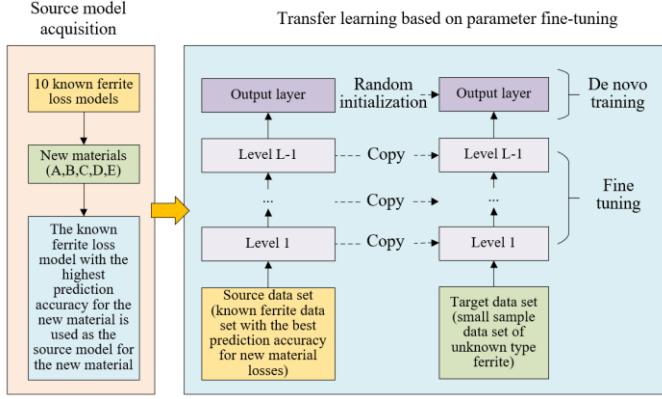


Fig. 4 Prediction of ferrite loss of unknown type based on parametric fine-tuning TL

In transfer learning, the input layer has a feature dimension size of 16. Firstly, a CNN module consisting of a convolutional layer, ReLU activation function, and max pooling layer is used for feature extraction. The 1D convolution performs feature extraction with 1 input channel, 16 output channels, a kernel size of 25, a pooling window size of 10, and a stride of 10 to extract and integrate multiple batches of time series features. The LSTM layer contains a bidirectional LSTM layer with an input size of 16, a hidden size of 32, and a layer number of 1. The output of the LSTM layer is used as the input to the fully connected layer. Fully connected layer 1 consists of three linear layers with a structure of 32-32-13 and ReLU activation function. Fully connected layer 2 consists of three linear layers with a structure of 16-16-1 and ReLU activation function. The optimizer selected is Adam optimizer with a learning rate of 0.0001. The special feature of the optimizer is that it uses the filter function to select the parameters with requires 'grad=True' in the model for optimization. The total number of parameters in the model is 7154.

IV. CONCLUSION

This paper proposed a neural network based core loss model for ferrite. The CNN+LSTM network is proposed to model the core loss under large core loss datasets. The transfer learning method is utilized to model the core loss under small core loss datasets. The accuracy of these neural network models are high.

REFERENCES

- [1] C. P. Steinmetz, "On the law of hysteresis," *Proc. IEEE*, vol. 72, no. 2, pp. 197–221, Feb. 1984.
- [2] 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," *IEEE Trans. Ind. Appl.*, vol. 37, no. 4, pp. 1055–1061, Jul./Aug. 2001.
- [3] K. Venkatachalam, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," *Proc. IEEE Workshop Comput. Power Electron.*, Jun. 2002, pp. 36–41.
- [4] W. Shen, F. Wang, D. Boroyevich, and C. W. Tipton, "Loss characterization and calculation of nanocrystalline cores for high-frequency magnetics applications," *IEEE Trans. Power Electron.*, vol. 23, no. 1, pp. 475–484, Jan. 2008.
- [5] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core loss calculation for magnetic components employed in power electronic systems," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 964–973, Feb. 2012.
- [6] M. Mu, and F. C. Lee, "A new core loss model for rectangular AC voltages," in *Proc. Energy Convers. Congr. Expo. Conf.*, 2014, pp. 5214–5220.
- [7] C. R. Sullivan, J. H. Harris, and E. Herbert, "Core loss predictions for general PWM waveforms from a simplified set of measured data," in *Proc. 25th Annu. IEEE Appl. Power Electron. Conf. Expo.*, 2010, pp. 1048–1055.
- [8] S. Barg, K. Ammous, H. Mejibri, and A. Ammous, "An improved empirical formulation for magnetic core losses estimation under nonsinusoidal induction," *IEEE Trans. Power Electron.*, vol. 32, no. 3, pp. 2146–2154, Mar. 2017.
- [9] G. Bertotti, "General properties of power losses in soft ferromagnetic materials," *IEEE Trans. Magn.*, vol. MAG-24, no. 1, pp. 621–630, Jan. 1988.
- [10] D. C. Jiles and D. L. Atherton, "Theory of ferromagnetic hysteresis," *J. Magn. Magn. Mater.*, vol. 61, pp. 48–60, 1986.
- [11] Y. Bernard, E. Mendes, and F. Bouillault, "Dynamic hysteresis modeling based on Preisach model," *IEEE Trans. Magn.*, vol. 38, no. 2, pp. 885–888, Mar. 2002.
- [12] E. Dogariu, H. Li, S. Wang, M. Luo and M. Chen, "Transfer learning methods for data-driven magnetic core loss modeling," *IEEE Workshop on Control and Modeling of Power Electronics (COMPEL)*, 2021.
- [13] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen and M. Chen, "MagNet: A Machine Learning Framework for Magnetic Core Loss Modeling," in *Proc. IEEE 21st Workshop Control Model. Power Electron.*, 2020, pp. 1–8.
- [14] H. Li et al., "MagNet: An Open-Source Database for Data-Driven Magnetic Core Loss Modeling," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2022, pp. 588–595.
- [15] D. Serrano et al., "Neural Network as Datasheet: Modeling B-H Loops of Power Magnetics with Sequence-to-Sequence LSTM Encoder-Decoder Architecture," in *Proc. IEEE 23rd Workshop Control Model. Power Electron.*, 2022, pp. 1–8.
- [16] E. Dogariu, H. Li, D. Serrano López, S. Wang, M. Luo, and M. Chen, "Transfer Learning Methods for Magnetic Core Loss Modeling," in *Proc. IEEE 22nd Workshop Control Model. Power Electron.*, 2021, pp. 1–6.
- [17] D. Serrano et al., "Why MagNet: Quantifying the Complexity of Modeling Power Magnetic Material Characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 11, pp. 14292–14316, Nov. 2023.
- [18] H. Li et al., "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15829–15853, Dec. 2023.
- [19] H. Li, D. Serrano, S. Wang, and M. Chen, "MagNet-AI: Neural Network as Datasheet for Magnetics Modeling and Material Recommendation," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15854–15869, Dec. 2023.

MagNet Challenge 2023 Final Report

Chushan Li, *Member, IEEE*, Yinan Yao, *Student Member, IEEE*, Tianxiang Hu, *Student Member, IEEE*,
Lumeng Xu, Yiyi Wang, Sichen Wang

Abstract- Modeling of magnetic material loss is fundamental in magnetic components design, yet no fully satisfactory model has been proposed due to the complexity of underlying physics. Equation based model can not guarantee precise results across the entire application range. Data driven model is expected to surpass the limitations of equation-based method, but current research still faces challenges in terms of model size, accuracy, and the applicability range. Therefore, a physics informed neural network (PINN) model with smaller size, higher accuracy, wide applicability range and reduced data dependency is proposed. The improved general Steinmetz equation is incorporated into NN for mitigation of model reliance on extensive datasets. The architecture of the proposed model combines Gated Recurrent Unit and Feedforward Neural Network, enabling the seamless integration of waveform, temperature, and frequency information within a unified framework. The proposed model has compact size of 4285 parameters, and achieves 95-prct error of 8.35%, 4.11%, 8.78%, 25.70%, and 8.74% for new materials A, B, C, D, E. Notably, the model exhibits reduced data dependency, as it achieves comparable accuracy with only 25 data points, whereas normal NN model that doesn't incorporate physics information require more than 300 data points for acceptable performance.

I. INTRODUCTION

Magnetics components, which are transformers and inductors, are widely used in power electronics systems for energy storage, energy conversion and electrical isolation, etc. Typical magnetic components occupy large volume of the power electronics system and has high power loss, thus limiting the improvement of efficiency and power density of the whole system. Establishing an accurate magnetics loss model is crucial for effective material selection, thermal design, and performance prediction of magnetic components, which provides essential guidance to the overall system design and optimization.

Magnetic material has complicated physical mechanism, which is the first main difficulty of core loss modeling. Also, due to the frequency varying, temperature varying and dc bias varying characteristics of the high permeability magnetic material, the loss model is a highly nonlinear system, which adds difficulty to the core loss modeling. Furthermore, factors including frequency, temperature and dc bias are coupled and changing together in real conditions, therefore adding complexity to characterizing magnetics materials[1].

Currently, there is no universally applicable and accurate physical formula for magnetic loss. Models based on physical mechanisms are complex and demand extensive experimental data, making them impractical for real-world applications[2]. The most commonly used magnetic loss model is the

Steinmetz formula and its improved models, such as the modified Steinmetz formula (MSE), the generalized Steinmetz equation (GSE) and the improved GSE equation (iGSE)[3]. While equation-based magnetic loss modeling is easy to use, it cannot guarantee precise results across the entire application range.

Data-driven approaches have been considered as an effective means to surpass the limitations of equation-based magnetic loss modeling. Consequently, they have emerged as a prominent research focus in the field of magnetic loss modeling. The large scale magnetics material database - Magnet- has laid a solid foundation for data driven based magnetics material loss modeling[4]. [4]has proposed three types of neural network (NN) models for predicting core loss: scalar-to-scalar, sequence-to-scalar, and sequence-to-sequence models. The scalar-to-scalar NN model maps scalar inputs such as frequency, flux density, and duty ratio to a scalar output that represents core loss. The sequence-to-scalar NN model maps excitation waveforms to the core loss output. The sequence-to-sequence NN model maps flux density waveforms to magnetic field intensity waveforms and utilizes the B-H hysteresis loop to determine core loss. The scalar-to-scalar model has the smallest size, with a total of 1,594 parameters. However, it is only applicable for predicting losses under specific waveform shapes. The sequence-to-scalar model can handle different waveform shapes but does not incorporate temperature into the framework. On the other hand, the sequence-to-sequence model can handle all waveform shapes and temperature variations. However, it has the largest model size with 28,481 parameters, and the prediction bias in magnetic field intensity (H) can result in large core loss deviations. Existing research still has limitations in terms of model size, accuracy, and applicability range.

Therefore, a physics informed neural network (PINN) model with smaller size, higher accuracy, full applicability range and reduced data dependency is proposed. Firstly, an iGSE incorporated PINN model is established. By incorporating iGSE formula into the model, the acceleration of neural network training and mitigation of their reliance on extensive datasets have been achieved. Additionally, taking into account considerations of model size, accuracy, and applicability, we have chosen to adopt a neural network architecture that combines Gated Recurrent Unit (GRU) and Feedforward Neural Network (FNN). The GRU unit serves as an effective mechanism for processing sequential information, while the FNN unit is responsible for handling scalar information. This structure enables the integration of waveform, temperature, and frequency information within this

unified framework. To validate the effectiveness of our proposed approach, we have conducted repetitive experiments and evaluations. These evaluations include comparisons with PINN and normal NN, as well as the magnetics core loss NN model proposed by Princeton[4]. By considering various performance metrics, including accuracy, model size, and data size dependency, we can comprehensively assess the advantages and limitations of our proposed model.

II. MODELING OF PHYSICS INFORMED NEURAL NETWORK INCORPORATING IMPROVED GENERAL STEINMETZ EQUATION REQUIREMENTS

PINN has been widely used in the field of artificial intelligence, particularly in situations with limited data or when dealing with novel materials and operating condition. By integrating physical principles and constraints into the NN architecture, the PINN model leverages the available domain knowledge to guide the learning process. Therefore, PINN effectively improves the data utilization and thereby reducing the data dependency of neural networks.

The improved general Steinmetz equation (iGSE) is the most commonly used empirical formula for magnetic losses, which calculates core losses under arbitrary waveforms. The formula is shown in (1).

$$P_{IGSE} = \frac{k_i}{T} \int_0^T \left| \frac{1}{\Delta B} \frac{dB}{dt} \right| (\Delta B)^\beta dt$$

$$k_i = \frac{k}{(2\pi)^{\alpha-1} \int_0^{2\pi} |\cos \theta|^\alpha 2^{\beta-\alpha} d\theta} \quad (1)$$

Core loss of piecewise linear waveforms can be calculated as (2), in which ΔB_n and D_n are the change of magnetics flux density and duty cycle of each segment in a piecewise linear waveform. By calculating the integral in (2), the compact form for core loss of piecewise linear waveforms is shown in equation (3).

$$P_c = \frac{k_i}{T} \left[\int_0^{D_1 T} \left| \frac{1}{\Delta B_1} \frac{\Delta B_1}{D_1 T} \right|^\alpha (\Delta B_1)^\beta dt + \int_{D_1 T}^{(D_1+D_2)T} \left| \frac{1}{\Delta B_2} \frac{\Delta B_2}{D_2 T} \right|^\alpha (\Delta B_2)^\beta dt + \dots \right] \quad (2)$$

$$P_c = k_i \sum_n^\infty (\Delta B_n)^\beta D_n^{1-\alpha} f^\alpha \quad (3)$$

The logarithms of core loss for sinusoidal and piecewise linear waveforms are shown in equation (4) and (5) respectively.

$$\log P = \log k + \alpha \log f + \beta \log B \quad (4)$$

$$\log P = \log k_i + \alpha \log f + \log \left(\sum_n^\infty (\Delta B_n)^\beta D_n^{1-\alpha} \right) \quad (5)$$

Then gradient relationships between power loss P and frequency f , flux density B can be derived. Equation (6) and (7) shows gradient relationships between $\log P$, $\log f$, and $\log B$ for sinusoidal waveforms. For piecewise linear waveforms, the

gradient of $\log P$ with respective to $\log f$ is the same as Equation (6). Equation (8) shows gradient of $\log P$ with respective to $\log \Delta B_1$.

$$\frac{\partial \log P}{\partial \log f} = \alpha \quad (6)$$

$$\frac{\partial \log P}{\partial \log B} = \beta \quad (7)$$

$$\frac{\partial \log P}{\partial \log \Delta B_1} = \frac{D_1^{1-\alpha}}{D_1^{1-\alpha} (\Delta B_1)^\beta + D_2^{1-\alpha} (\Delta B_2)^\beta} \times \beta (\Delta B_1)^{\beta-1} \quad (8)$$

The iGSE formula is embedded in the neural network to assist in the learning process, as is shown in Fig. 1. The key is to use formula (6)-(8) to constrain the change of output with respect to the input, thereby determining the direction of parameter updates. The loss function of IGSE incorporated PINN is shown in equation (9)-(12). Equation (9) defines the residual1, which constrains the change of power loss with respective to frequency. Equation (10) and (11) defines the residual2 for sinusoidal and piecewise linear waveforms respectively, which constrains the change of power loss with respective to flux density. A smaller residual indicates that the updated direction of the output is more consistent with physical constraints.

$$residual_1 = \frac{\partial \log P}{\partial \log f} - \alpha \quad (9)$$

$$residual_2 = \frac{\partial \log P}{\partial \log B} - \beta \quad (10)$$

$$residual_2 = \frac{\partial \log P}{\partial \log \Delta B_1} - \frac{D_1^{1-\alpha} \beta (\Delta B_1)^{\beta-1}}{D_1^{1-\alpha} (\Delta B_1)^\beta + D_2^{1-\alpha} (\Delta B_2)^\beta} \quad (11)$$

$$loss = \frac{1}{N} \sum (\hat{P} - P)^2 + \frac{1}{N} residual_1^2 + \frac{1}{N} residual_2^2 \quad (12)$$

To verify the effectiveness of the PINN model, the normal NN model that has no embedded physics information is trained and tested for comparison. Repeating experiments are carried out while varying the amount of data for training. Fig. 2 shows the average relative errors of N87 material from MagNet Database, with different amount of data for training ranging from 25 to 1600. The error results are averaged over 10 trials to avoid randomness. The iGSE incorporated PINN has much lower relative error when the amount of data points available for training is small. This observation suggests that the integration of additional constraints from physical formulas in the PINN model facilitates faster learning and enhances the efficient utilization of data by the neural network. When the amount of data points for training reaches 1600, the PINN model and the normal NN model reaches the same average relative error at around 20%. As the training data becomes more abundant, it is possible for the performance of PINN to slightly lag behind that of traditional neural networks. This can be attributed to the errors introduced by incorporating physical constraints into the model.

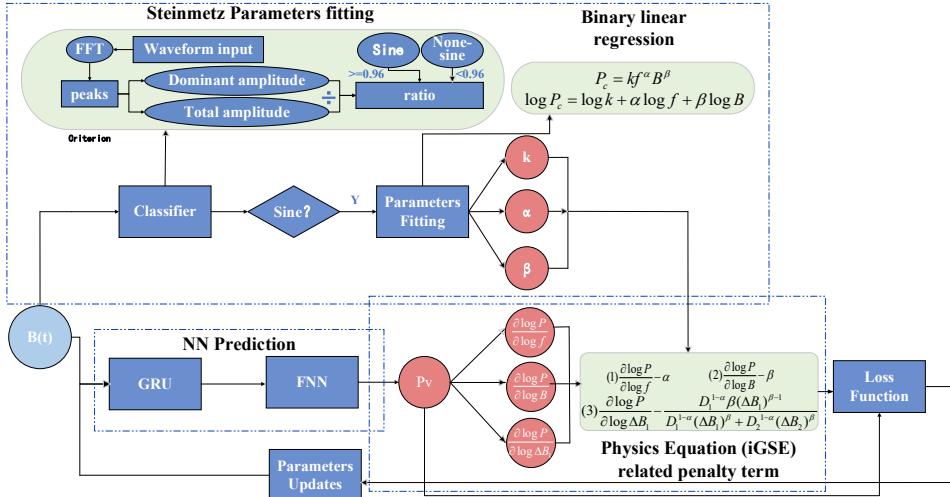


Fig. 1 IGSE incorporated PINN

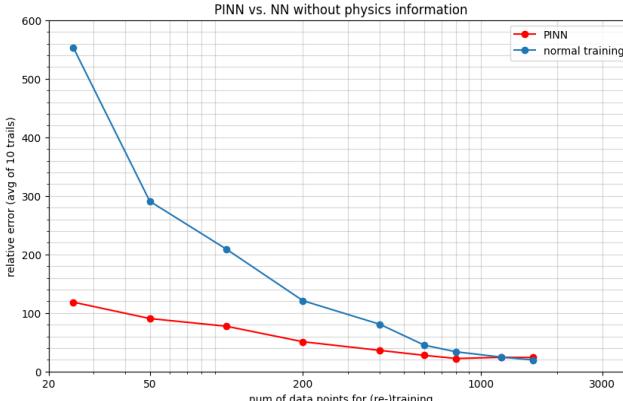


Fig. 2 Relative error of PINN and normal NN with a varied amount of training data

III. UNIFIED NN FRAMEWORK WITH SMALLER SIZE, HIGHER ACCURACY AND WIDE APPLICABILITY RANGES

The tradeoff between model size and accuracy is a crucial challenge in NN modeling, and an overall characterization of magnetics material loss behavior requires a unified framework. Therefore, a novel structure that enables the seamless integration of waveform, temperature, and frequency information should be constructed with compact size and high accuracy.

The proposed model should be able to handle sequential input flux density waveform $B(t)$, as well as multiple scalar inputs. The loss of magnetic materials is significantly affected by temperature, making it crucial to model the influence of temperature on the output loss accurately. To capture the temperature-dependent behavior of magnetic material loss, temperature T serves as a vital input for the proposed model. To incorporate physical information into the neural network, as mentioned in section 2, the proposed model utilizes formulas (9)-(12). This requires $\log f$, $\log B$ and the magnetic flux density variation value of the first segment of the

piecewise linear waveform ΔB_1 to be fed into the NN model as well. In summary, the proposed model should have two stages, to process sequential information $B(t)$, as well as scalar inputs T , $\log f$, $\log B$, and ΔB_1 .

GRU and LSTM are two widely used variants of recurrent neural networks (RNNs) that incorporates gating mechanisms, which renders the ability to handle time-series information. Compared to LSTM, GRU features a simpler gating mechanism comprising only an update gate and a reset gate. Consequently, GRU boasts fewer parameters and exhibits faster convergence. Hence, we opted for GRU as the network architecture for waveform processing. The GRU stage has 32 cell states and 32 hidden states. It takes the flux density waveform $B(t)$ as input, and the outputs are fed into FNN for further processing.

The FNN stage consists of one input layer, five hidden layers and one output layer that generates power loss prediction. The FNN aggregates the outputs of GRU stage and other four inputs, which are T , $\log f$, $\log B$ and ΔB_1 . The proposed model contains 4285 parameters in total. The overall structure of our proposed model is shown in Fig. 3.

The model is trained and tested under all types of waveforms and at all temperature conditions for N87 material, which contains 40616 data points in total. The prediction results have an average error of 1.36%, RMS error of 1.83%, and maximum error of 17.8%. The error distribution is shown in Fig. 4.

The same training process are also conducted on other materials provided in MagNet Database, including 3C90, 3C94, N27, etc., to verify the effectiveness of the proposed model. Fig. 5 shows the error histograms of 3C90 and 3C94, which contains 40713 and 40068 data points respectively.

Comparisons with the proposed PINN model and other existing models are summarized in TABLE I. It can be concluded that the proposed model reaches higher accuracy with a compact model size and wide applicable range.

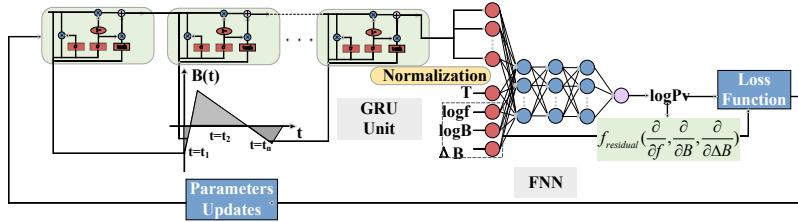


Fig. 4 The network structure of GRU based magnetic core loss modeling

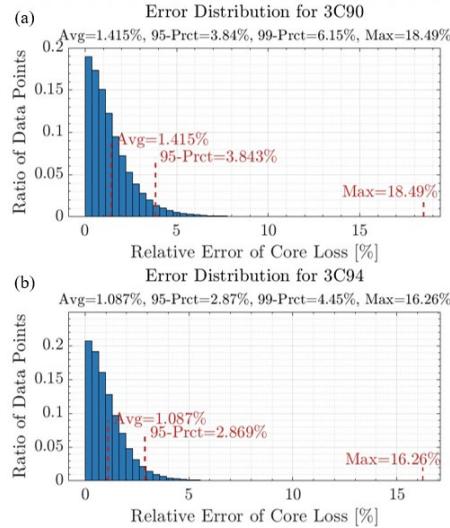


Fig. 5 (a) Error histograms of 3C90 (b) Error histograms of 3C94

IV. MODEL PERFORMANCE EVALUATION AND FINAL TEST RESULTS

To provide invaluable insights into a widely applicable training methodology for magnetics material modeling and deliver a meticulous evaluation of the model we put forth, we delved into the intricate technical aspects of the model training process and model evaluation. Additionally, exploration of the amount of data required for new material training is conducted as important part of our analysis. Finally, the 5 new materials are trained, and the evaluation of our proposed model in terms of mode size and accuracy are summarized.

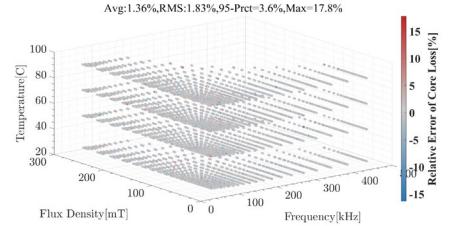


Fig. 3 Error distribution of N87 material prediction results under all types of waveforms and all temperature conditions

A. Data Pre-processing and Model Training Method

Data pre-processing is the first and fundamental step for model training. For our proposed PINN model, it can be decomposed into mainly three steps:

(1) Sequence and scalar inputs and outputs processing: To minimize the impact of phase, and enhance the model's resilience to noise, a circular random phase shift is done on each magnetic density waveform, complemented by superposition with white noise. And frequency information is incorporated in the waveform by down sampling. To enhance the clarity of the input-output relationships, the output power loss, and input flux density, and frequency are logarithmically transformed.

(2) Classification of waveforms: This classification was accomplished by employing Fast Fourier Transform (FFT) to discern frequency components. Three waveform class—sinusoidal, triangular, and trapezoidal—are categorized and the change of magnetics flux ΔB , the duty cycles of each segment of piecewise linear waveforms D_n are recorded for later incorporation of formulas (9)-(12).

(3) Steinmetz parameters fitting: the classifier described in step (2) then passes the sinusoidal waveforms to the Steinmetz parameters fitting function, which employs the least squares method. This fitting process aims to estimate the Steinmetz parameters that characterize the magnetic properties of the material under consideration.

 TABLE I
COMPARISONS WITH THE PROPOSED PINN MODEL AND OTHER EXISTING MODELS

Model	Model Size	Applicable Range	Test Condition		Error (%)
			Waveform Shape	Temperature	
Sequence-scalar[4]	5569	Handle only one temperature	tri.,	25°C	Avg:2.09% Max:14.35%
Sequence-sequence[4]	28481	Handle all waveforms and all temperature	Sine, tri., trape.	25°C,50°C, 70°C,90°C,	Avg: 4.2% Max: 30%
Proposed PINN model	4285	Handle all waveforms and all temperature	Sine, tri., trape.	25°C,50°C, 70°C,90°C,	Avg:1.36% Max:17.8%

Due to the limited training dataset available for new materials, we applied k-fold training to maximize the utilization of the dataset for neural network training. The training dataset is randomly split into training and test set as 80% and 20%. Then use k-fold training to further divide the first part into k subsets, with one subset serving as the validation set and the remaining k-1 subsets used for training during each iteration. Under the considerations of available data quantity and computational resources, the decision has been made to use k=5 for training. Therefore, thorough evaluation on the model's performance can be conducted and the most suitable model that achieves the best overall performance across the five folds can be chosen.

MagNet database has provided abundant magnetics loss data for materials including N87, N49, 3C90, etc, which provides basis for new material training. During the initial stages of the competition, we trained the magnetic loss model using a substantial amount of data, which serves as a valuable starting point for training new materials. In order to assess the model's ability to generalize, we conducted zero-shot experiments by directly testing the pre-trained N87 material loss model on material A. The results demonstrated that the model captured some knowledge about the loss mechanism and the factors that influence it, while due to the inherent differences between materials, the zero-shot errors were large and further training needs to be conducted. Subsequently, normal training and transfer learning are conducted. Transfer learning uses the pretrained model as a start point and retrains with the new material data, while normal training uses randomly initialized neural network. Transfer training provides better accuracy results, which verifies the effectiveness of the pre-trained model as a starting point for training new materials. leveraging the knowledge acquired from the previous training. TABLE II summarizes the zero-shot, normal training, and transfer learning results on material A, testing on a same test set and the error results are averaged over 10 trials to avoid randomness and for better comparison.

B. Data Dependency Exploration

The neural network performance dependency on the amount of data are explored, to give valuable insights into determining the minimum amount of data required to achieve an acceptable error level during training. The figure shows the relative error results of material A with varying amount of data for training. The proposed PINN model is compared with normal NN model that doesn't incorporate physics formula. Repetitive experiments are conducted with different amount of data for training ranging from 25 to 2400, and the error results are averaged over 10 trials to avoid randomness.

TABLE II

ZERO-SHOT, NORMAL TRAINING AND TRANSFER LEARNING RESULTS OF MATERIAL A

Error	Zero-shot	Normal training	Transfer learning
Avg. Relative Error	35.89	10.52	7.46
RMS Relative Error	41.89	15.50	10.77
Max Relative Error	137.08	116.12	85.61

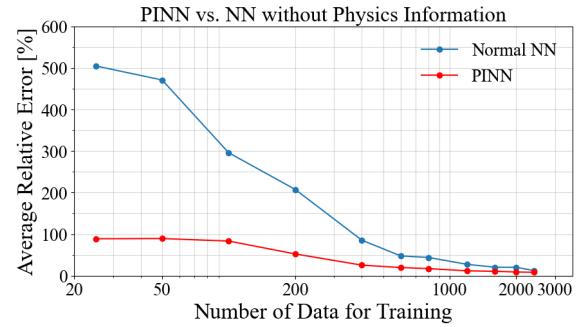


Fig. 6 Data dependency results of normal NN and PINN for material A

Remarkably, the model exhibits reduced data dependency, as it achieves comparable accuracy with just 25 data points, whereas normal NN model that doesn't incorporate physics information require more than 300 data points for acceptable performance. The results show the proposed PINN model helps enhance the effective utilization of data by neural networks. These findings provide crucial insights into determining the optimal amount of data necessary for training to achieve an acceptable error level. By understanding the data dependency, researchers and practitioners can make informed decisions regarding the dataset size and allocation, ultimately enhancing the overall performance and efficiency of the neural network model.

C. Final Evaluation Results

Using our proposed model, we successfully obtained the final results for the 5 new materials. The model size and accuracies are summarized in TABLE III.

The model demonstrated promising accuracy for Materials A, B, C, and E, with relatively low average and 95% percentile errors. However, the accuracy for Material D was comparatively lower, with higher errors.

To improve the accuracy for Material D and enhance overall performance, further analysis is needed to identify the specific reasons behind the lower accuracy. Possible improvements could include acquiring additional data for Material D, refining the data pre-processing steps, or exploring alternative model architectures.

In conclusion, our study developed a training methodology for magnetic material modeling and evaluated the proposed model on 5 new materials. The results showcased the effectiveness of our approach, although further enhancements are required for certain materials. The insights gained from data dependency analysis can guide researchers in determining optimal data sizes for training, leading to improved neural network performance and efficiency.

TABLE III
TEST RESULTS OF 5 NEW MATERIALS

Material	Model Size	Accuracy
Material A	4285	Avg:2.90%, 95-prct: 8.08%
Material B		Avg:1.47%, 95-prct: 4.04%
Material C		Avg:3.17%, 95-prct: 8.44%
Material D		Avg:8.04%, 95-prct: 23.27%
Material E		Avg:2.99%, 95-prct: 8.85%

REFERENCES

- [1] McLyman C W T. Transformer and inductor design handbook[M]. CRC press, 2016.
- [2] Mühlethaler J. Modeling and multi-objective optimization of inductive power components[D]. ETH Zurich, 2012.
- [3] Z. Li, W. Han, Z. Xin, Q. Liu, J. Chen and P. C. Loh, "A Review of Magnetic Core Materials, Core Loss Modeling and Measurements in High-Power High-Frequency Transformers," in CPSS Transactions on Power Electronics and Applications, vol. 7, no. 4, pp. 359-373, December 2022, doi: 10.24295/CPSSTPEA.2022.00033.
- [4] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen and M. Chen, "MagNet: A Machine Learning Framework for Magnetic Core Loss Modeling," 2020 IEEE 21st Workshop on Control and Modeling for Power Electronics (COMPEL), Aalborg, Denmark, 2020, pp. 1-8, doi: 10.1109/COMPEL49091.2020.9265869.