MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

A BERT-Based Surrogate Model for Permanent Magnet Motor Cogging Torque Prediction

Sun, Siyuan; Wang, Ye; Koike-Akino, Toshiaki; Yamamoto, Tatsuya; Sakamoto, Yusuke; Wang, Bingnan

TR2026-123 September 09, 2026

Abstract

Motor performances such as cogging torque and torque ripple are difficult to predict accurately with surrogate models. In this work, we propose text-transformer-based models to tackle the problem. We first formatted the motor design parameters into natural-language sentences describing the design, which are fed into a text transformer model, and fine-tuned on the motor design dataset to make predictions. Compared with conventional artificial neural network models, our approach is more flexible in the input parameter dimensions, and can be jointly trained on multiple motor types. Numerical tests are performed on two datasets with different types of interior permanent magnet motor designs, and initial results show substantial improvements on cogging torque prediction accuracy over artificial neural network models on both datasets.

International Conference on Electrical Machines (ICEM) 2026

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

A BERT-Based Surrogate Model for Permanent Magnet Motor Cogging Torque Prediction

Siyuan Sun
Department of Computer Science, Iowa State University,
Ames, IA 50011, USA

Ye Wang
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139 USA

Toshiaki Koike-Akino
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139 USA

Tatsuya Yamamoto
AI Transformation Innovation Center Mitsubishi Electric Corporation
Kamakura, Kanagawa 247-8501 Japan

Yusuke Sakamoto
Advanced Technology R&D Center Mitsubishi Electric Corporation
Amagasaki, Hyogo 661-8661 Japan

Bingnan Wang
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139 USA

Abstract: Motor performances such as cogging torque and torque ripple are difficult to predict accurately with surrogate models. In this work, we propose text-transformer-based models to tackle the problem. We first formatted the motor design parameters into natural-language sentences describing the design, which are fed into a text transformer model, and fine-tuned on the motor design dataset to make predictions. Compared with conventional artificial neural network models, our approach is more flexible in the input parameter dimensions, and can be jointly trained on multiple motor types. Numerical tests are performed on two datasets with different types of interior permanent magnet motor designs, and initial results show substantial improvements on cogging torque prediction accuracy over artificial neural network models on both datasets.

***Index Terms*—Electric Motors, Surrogate Model, Machine Learning, Text Transformer.**

I. Introduction

Electric machines are widely used in households and various industries, and their technologies and design principles are well established. However, the requirements for motor design and customization, especially for new applications such as transportation electrification and factory automation, always pose new challenges to motor designers. Parameter sweeping or iterative optimization methods are often utilized to evaluate a large number of design candidates before identifying the optimal design for a specific task. The accurate analysis of a motor design typically relies on numerical simulations based on finite-element analysis (FEA), which are time-consuming, especially when various operating points are evaluated for one design. Surrogate model based optimization has been investigated to speed up the process [1]. Due to the highly nonlinear nature, the accuracy of conventional surrogate models is not sufficient for the prediction of certain motor performances such as torque profile and efficiency map. In recent years, machine learning and deep learning methods have been developed and applied to many applications including motor design [2], [3], due to their powerful capability to emulate highly nonlinear functions. For example, several studies [4]–[7] have proposed neural network-based approaches to address various critical tasks, such as generating innovative motor designs, predicting physical responses for given designs, and detecting potential faults during operations.

Currently, most existing machine-learning-based surrogate modeling methods represent motor designs either as (i) a vector of geometric parameters or (ii) a colored 2D cross-sectional image. Different models can be constructed with either parameter-based inputs or image-based inputs. Deep learning models such as convolutional neural networks (CNNs) and vision transformers utilizing image-based inputs can achieve a higher degree of predictive accuracy for more complicated tasks, while also require larger dataset and longer training time.

For parameter representations, researchers typically employ artificial neural network (ANN) models as surrogate models to predict the physical responses of a motor design. However, such networks typically accept only numerical inputs, which strips away the contextual and semantic richness of the parameters. In addition, a motor design generally contains only a limited number of input parameters, which carry a limited amount of information. As a result, when relying solely on numerical representations, the model may fail to capture sufficient information, significantly restricting its performance.

Natural language processing has seen remarkable advances in recent years. Among these, the Bidirectional Encoder Representations from Transformers (BERT) [8] model stands out as a foundational architecture for handling natural language inputs. Unlike earlier unidirectional models that process text sequentially and consider only preceding context, BERT adopts a bidirectional approach through its Transformer-based architecture. By attending to both the preceding and succeeding words simultaneously, BERT can better capture the full contextual meaning of each word. This bidirectional context

This work was done when S. Sun was with MERL as research intern. Corresponding Author: Bingnan Wang (bwang@merl.com).

modeling allows BERT to resolve ambiguities and understand nuanced language patterns more effectively. As a result, BERT demonstrates significant improvements across a wide range of natural language understanding tasks, such as sentiment analysis [9], [10], information retrieval [11]–[13] and language translation [14].

In this paper, we propose a natural-language format based on the numerical parameters of a motor design that is easy to generate and convey key design information. We then introduce a novel BERT-based surrogate model to more accurately and efficiently estimate the physical responses of motor designs, advancing the application of modern deep learning techniques in motor design optimization. We demonstrate that BERT models, combined with our proposed input format, significantly outperform traditional deep ANN models in terms of accuracy, as shown by the cogging torque prediction of two types of interior permanent magnet synchronous motors (IPMSMs). In addition, compared with ANN models, the proposed framework is easy to adapt to different motor design topologies with different input parameters. This advantage is demonstrated by the joint training on the two types of IPMSM.

II. Problem Definition

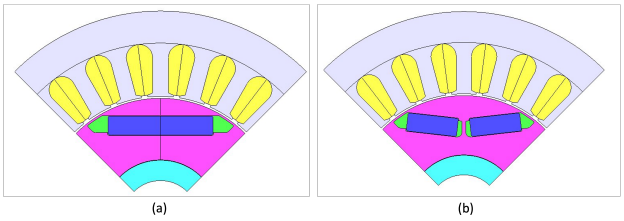

Fig. 1. Example design of IPM motors with (a) flat magnets and (b) V-shape magnets.

In this work, we built datasets to evaluate the cogging torque of 4-pole 24-slot IPMSMs for two types of magnet configurations. Fig. 1 shows a representative design of one magnetic pole for each configuration. All the motor designs in the two datasets have the same inner and outer dimensions. Specific magnetic design parameters play a crucial role in determining a motor's performances. Key design parameters for the two types of IPM motors are indicated in Fig. 2. For motors with flat magnets, a total of 19,375 designs are generated by sampling 13 geometrical parameter values from predefined ranges, and evaluated with FEA simulations using JMAG. The list of design parameters for flat magnet designs corresponding to Fig. 2 is shown in Table I. For motors with V-shape magnets, a total of 35,001 designs are generated by adjusting 17 geometrical parameters, and evaluated with FEA simulations. The list of design parameters for V-shape magnet designs corresponding to Fig. 2 is shown in Table II.

The simulated torque waveform for each design is decomposed into a Fourier series including the dominating harmonic

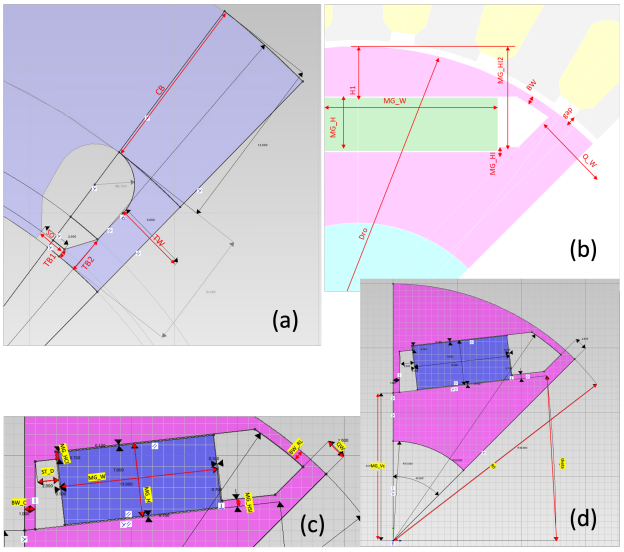

Fig. 2. Design parameters for the two types of IPM motors. (a) stator design parameters for both motor types, (b) rotor design parameters for the motors with flat magnets, (c) and (d) indicate design parameters for the motors with V-shape magnets.

Parameter Name	Parameter Description
BW	Width of rotor bridge
CB	Stator core back iron length
Dro	Rotor outer radius
H1	Distance of magnet to rotor surface
MG_H	Magnet height
MG_HI	Height of magnet slot bottom wedge
MG_W	Magnet width
Q_W	Distance of flux barrier to Q axis
SO	Stator slot opening width
TB1	Stator slot opening height
TB2	Stator slot wedge height
TW	Spacing between neighboring stator slots
gap	Air gap length

TABLE I. Flat Design Parameters and Their Corresponding Values

terms:

$$T(\theta) = A_{12}\cos(12\theta) + B_{12}\sin(12\theta) + A_{24}\cos(24\theta) + B_{24}\sin(24\theta). \quad (1)$$

The four Fourier coefficients $A_{12}, B_{12}, A_{24}, B_{24}$ are recorded to recover the cogging torque. Higher order harmonics are neglected, as this reconstruction has been shown to be accurate, with errors within a few percent. Each entry in the compiled dataset includes the value of design parameters, the corresponding 2D cross section image, and four Fourier coefficients. In addition, the induced voltage EMF is also recorded.

Surrogate models are trained to predict the Fourier coefficients for a given motor design, which can be used to recover the torque waveform according to Eq.(1). The cogging torque is then determined by the difference between maximum and minimum value of the torque: $T_c = \max(T(\theta)) - \min(T(\theta))$. Based on previous research [15], the performance of cogging torque prediction with the Fourier series approach is generally better than directly predicting cogging torque value using

Parameter Name	Description
BW	Width of the rotor bridge
CB	Stator core back iron length
MG_H	Magnet height
MG_W	Magnet width
QW	Distance of flux barrier to Q axis
R0	Rotor outer radius
MG_HSI	Height of magnet slot bottom wedge
theta	Magnet angle
MG_HCI	Height of magnet slot top wedge
ST_D	Inner flux barrier width
BW_C	Central bridge width
MG_Vc	Radius of the V-shape
gap	Air gap length
TW	Spacing between neighboring stator slots
SO	Stator slot opening width
TB1	Stator slot opening width
TB2	Stator slot wedge height

TABLE II. V-shape Design Parameters and Their Corresponding Values

the same model. We aim to develop models that predict the cogging torque, with the objective of reducing the root-mean-square error (RMSE) of the prediction.

III. Proposed Methodology

In this section, we present the proposed surrogate modeling methodology: We first represent a motor design with natural language descriptions based on the given list of parameters, and then feed that into a text transformer model to make the predictions.

A. Natural Language Description of Motor Design

As shown in Fig. 2 and Tables I and II, the two types of IPMSMs have the same design parameters on the stator side, and distinct parameters for the rotor designs specific to each type. If we represent these designs using only fixed-length parameter arrays, one with 13 elements for motors with flat magnets (Motor A), and the other with 17 for motors with V-shape magnets (Motor B), all contextual and semantic information beyond the raw numbers is lost. This indicates the limitations of using purely numerical input representations.

To address this issue, we propose extending the regular numerical representation into a natural language description. Each description consists of two parts: the **model** name and the **model parameter description**. The model name is used to distinguish between different motor types, while the parameter description maps each parameter name to its corresponding numerical value. In Figure 3, we present both the numerical values and the corresponding natural language descriptions for two different motor designs. This comparison highlights the contrast between the two types of input representations. Using only numerical data, we can observe that Motor A has 13 input parameters, while Motor B has 17. However, when these inputs are expanded into natural language descriptions, we gain a deeper understanding of the designs. The natural language descriptions reveal that the two motors follow different design principles. And it also clarifies the meaning of each numerical value. These additional details can help the model better capture the underlying features and characteristics of each motor design.

B. Text Transformer Model

Traditional ANNs are not well-suited to processing natural language inputs. To address this limitation, we propose employing a new type of neural network architecture that is capable of handling and interpreting natural language descriptions effectively. In particular, BERT-based models [8] have demonstrated strong performance across a wide range of natural language processing (NLP) tasks. The core architecture of BERT consists of two main components: a tokenizer and an encoder. In this section, we introduce the BERT model and explain how it can be integrated with motor design applications.

1) Tokenizer: Before feeding our description into the BERT model, we must first transform the raw text into tokens that the model can understand. This pre-processing step is typically handled by a tokenizer layer. It involves the following steps:

- Split words into sub-word units using a method such as WordPiece [16], which helps the model handle rare or combined words more effectively.
- Insert special tokens required by BERT. In our case, [SEP] will be used to distinguish between different sequences. [PAD] is used to adjust sequence length for batch processing.
- Convert each word (or sub-word) in the sequence into a unique integer ID.

The final token list will serve as the input to the BERT model, with each BERT model utilizing its own tokenizer to ensure optimal performance.

2) BERT-based encoder: Compared with other sequence-based models such as recurrent neural networks (RNNs), One of the key benefits of BERT is its "bidirectional" processing, enabled by the self-attention mechanism in transformer-based models. An attention layer takes three vectors as input:

- **Query**(Q): Target Word
- **Key**(K): The information that each token "offers" to others.
- **Value**(V): Actual information that will be passed as output

In NLP, we are especially interested in the relationships between words in a sentence. Therefore, the Query Q, Key K, and Value V vectors are all derived from the same input sentence, a setup known as self-attention. To implement the self-attention mechanism, we define three learnable weight matrices: W_Q, W_K, and W_V. These are used to transform the input X into the corresponding representations:

$$Q = XW_Q, \qquad K = XW_K, \qquad V = XW_V. \tag{2}$$

The self-attention layer computes an attention score S based on the Query Q and Key K. It then uses this score to determine how much of each Value V should contribute to the final output. This process is formally described by the following equations:

$$M = \text{Softmax}(\frac{QK^\top}{\sqrt{d_K}}), \qquad \text{Output} = MV \tag{3}$$

Motor A Numerical Input: [0.7, 5, 78, 7, 6, 0.4, 34, 12, 1, 0.8, 1.2, 4.5, 0.3]

Motor A Nature Language Description:
[Motor Design Name]This is a Flat interior permanent magnet motor design.
[Motor Design Parameter Description] This design can be described with the following parameters in millimeters: Width of rotor bridge is 0.7; Stator core back iron length is 5.0; Rotor outer radius is 78.0; Distance of magnet to rotor surface is 7.0; Magnet height is 6.0; Height of magnet slot bottom wedge is 0.4; Magnet width is 34.0; Distance of flux barrier to Q axis is 12.0; Stator slot opening width is 1.0; Stator slot opening height is 0.8; Stator slot wedge height is 1.2; Spacing between neighboring stator slots is 4.5; Air gap length is 0.3.

Motor B Numerical Input: [38, 0.6, 1.5, 6, 10, 0.8, 5, 0.2, 5.5, 1.5, 21, 0.9, 3.5, 7, 2.8, 1.6, 0.8]

Motor B Nature Language Description:
[Motor Design Name] This is an Vshape interior permanent magnet motor design.
[Motor Design Parameter Description] This design can be described with the following parameters in millimeters: Rotor outer radius is 38.0; Width of rotor bridge is 0.6; Distance of flux barrier to Q axis is 1.5; Magnet height is 6.0; Magnet width is 10.0; Height of magnet slot bottom wedge is 0.8; Magnet angle is 5.0; Height of magnet slot top wedge is 0.2; Inner flux barrier width is 5.5; Central bridge width is 1.5; Radius of the V is 21.0; Air gap length is 0.9; Spacing between neighboring stator slots is 3.5; Stator core back iron length is 7.0; Stator slot opening width is 2.8; Stator slot opening height is 1.6; Stator slot wedge height is 0.8.

Fig. 3. Natural language example for motor design.

Here, d_K is the dimensionality of the Key vector. The division by d_K serves to scale down the raw dot product scores, preventing them from becoming too large, which stabilizes the softmax output. The softmax function then normalizes the scores into a probability distribution, ensuring the attention weights sum to 1.

This unique self-attention mechanism enables the BERT-based encoder to process each word using context from both directions, allowing it to capture richer and more comprehensive information compared to previous unidirectional methods.

Before applying a BERT model to a specific downstream task, it is typically pre-trained on a large corpus of text to develop a general understanding of language, including how words relate to one another within and across sentences. This pre-training phase enables BERT to learn rich linguistic features such as grammar, syntax, and semantics. As a result, the model requires significantly less labeled data during fine-tuning for downstream tasks, since it already possesses a strong foundational knowledge of language structure.

C. Text Transformer for Motor Design

Given the strong performance of BERT-based models in NLP, we propose to integrate BERT with our natural language description framework for motor design to predict the performance of a motor.

In our approach, the natural language description of a motor design is first passed through a tokenizer, which converts the raw text into a sequence of tokens. This token sequence is then fed into a BERT-based encoder, which generates a contextualized representation of the input. Finally, a multi-layer perceptron (MLP) is used as a regressor to map the encoded representation to the predicted physical response of the motor. The detailed process is described in Figure 4.

This setting provides several benefits compared with previous parameter-based methods on motor design task.

- Easy to generate inputs: Compared to previous parameter-based approaches, our proposed input format is easier to generate. We simply define a template and fill in the parameters using natural language.
- More freedom for input: Traditional ANN models accept only numerical inputs, which restricts the types of information that can be incorporated. In contrast, our proposed method allows for greater flexibility by enabling the inclusion of any relevant information in natural language form.
- Adapt to different designs easily: When working with multiple motor designs that have different numbers of parameters, our method remains adaptable since the input length does not need to be fixed. In contrast, traditional artificial neural networks (ANNs) typically require inputs of a fixed size.
- Support higher model complexity: Parameter-only inputs limit the richness of information that can be provided. By leveraging natural language descriptions, our model captures more context and allows for the use of more complex architectures.

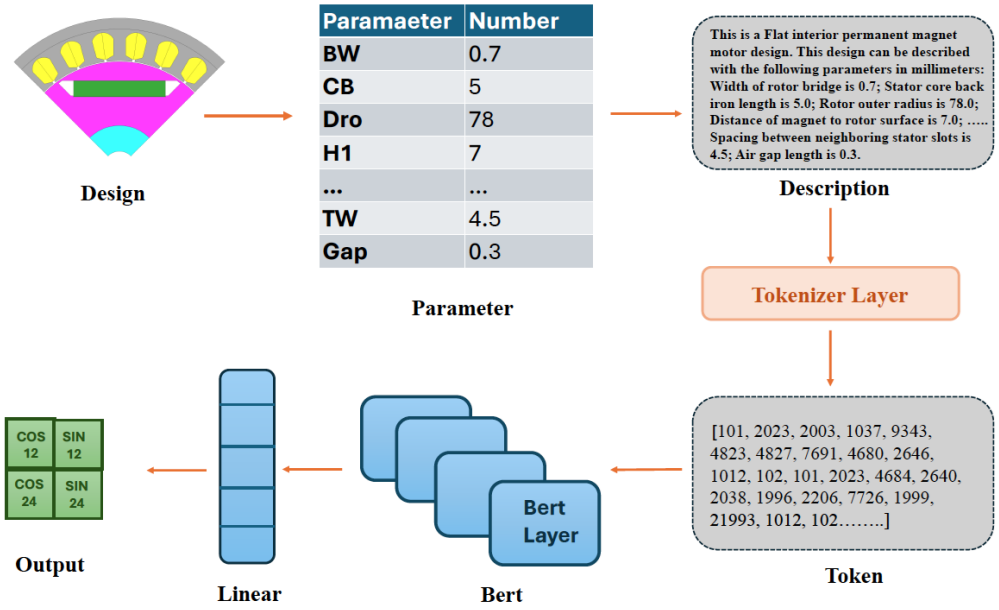

Fig. 4. Illustration for proposed text transformer with natural language description as input.

IV. Numerical Tests & Results

To evaluate the performance of BERT-based models utilizing natural language descriptions, we compared them against traditional ANNs that process numerical inputs.

A. Models and Hyperparameters

Model	Parameter Number
ANN-2-layer	10000 to 30000
ANN-3-layer	11000 to 4M
ANN-4-layer	31000 to 12M
ResNet101	42.8M
ResNet152	58.5M
DistilBERT	66M
BERT	110M

TABLE III. Parameter Number for each network. The parameter number of ANN model refers to one single model.

The network scales of model used in our evaluation are described in table III. Specifically, we implemented 2-layer, 3-layer, and 4-layer ANN architectures. To ensure optimal performance for each ANN model, we conducted hyperparameter tuning on key parameters, including network width, learning rate, number of training epochs, and weight decay coefficient. This tuning process aimed to identify the most effective training configurations for each model. In addition, we train one ANN model for each output parameter; therefore, 4 ANN models are used to predict the 4 Fourier series coefficients to reconstruct the cogging torque. In comparison, for larger models including ResNet and BERT variants, only one model is trained to predict all 4 Fourier coefficients.

For BERT-based models, we evaluate the following two variants:

- DistilBERT [17]: A lightweight model distilled from the full-size BERT.
- BERT [8]: The original full-size BERT model, pre-trained on a broad corpus of general-domain texts.

DistilBERT is a lightweight model with approximately 66 million parameters. In contrast, BERT has around 110 million parameters. Both DistilBert and BERT are pre-trained on general-domain texts.

Each IPM dataset is split into training (70%), validation (10%), and test (20%) datasets. For comprehensive evaluation, all six models introduced in Section IV-A will be compared, where the input for ANNs is numerical parameter values, and the input for BERT-based models is the natural language description introduced in III-A. For all three Text Transformer models, we fine-tune them on each IPM motor dataset over 300 epochs. Given that this is a fine-tuning setting, we adopt a small learning rate of 5×10^{-5} to ensure stable convergence.

B. Results

The prediction results are evaluated using RMSE on the test dataset. Specifically, the models predict the four Fourier transform coefficients, which are then used to reconstruct the coggin torque according to Equation 1. The RMSE is computed between the reconstructed cogging torque and the ground-truth in the test dataset. The corresponding results are presented in Table IV. The text transformer (BERT-based

Model	Flat	V-shape
ANN-2-layer	0.780 ± 0.607	0.390 ± 0.330
ANN-3-layer	0.649 ± 0.554	0.258 ± 0.242
ANN-4-layer	0.521 ± 0.455	0.389 ± 0.338
ResNet-101	1.010 ± 0.904	0.299 ± 0.283
ResNet-152	0.995 ± 0.897	0.299 ± 0.286
DistilBERT-cased	**0.309** ± 0.231	0.134 ± 0.109
BERT-cased	0.528 ± 0.425	**0.125** ± 0.101

TABLE IV. Performance Comparison for predicting cogging torque on IPMSM test datasets with Flat and V-shape magnets, with RMSE as evaluation metric.

Model	Cogging Torque
DistilBERT	0.284 ± 0.232
BERT	0.335 ± 0.289

TABLE V. Performance Comparison for cogging torque on joint dataset (both Flat and V-shape magnets) on two BERT models, with RMSE as evaluation metric.

model) captures the distribution of the cosine and sine coefficients of cogging torque better than the ANN model. As a result, for the Flat design, the best RMSE achieved by the BERT-based model is 0.309 (DistilBert), while the best ANN model can only reach 0.521 and best ResNet model can only reach to 0.995. Interestingly, the full BERT model does not outperform DistilBERT on the Flat dataset, suggesting that the smaller model may be sufficient under the current data and regularization settings. For the V-shape design, a similar trend is observed: the best RMSE from the BERT-based model is 0.125, compared to 0.258 for the ANN model and 0.299 for the ResNet Model.

Moreover, we observe noticeable differences among the various BERT-based model variants. When the amount of training data is limited, smaller models such as DistilBERT are often sufficient to achieve competitive performance, offering a good balance between efficiency and accuracy. In contrast, when ample training data is available, both small and large models are capable of delivering strong results. In such cases, the choice between model sizes may depend more on computational resources and inference speed requirements rather than performance alone.

Another important observation is that the performance improvement is not solely due to the increased number of tunable parameters. DistilBERT has a comparable number of parameters to the ResNet variants, as well as to the combined total of the four 4-layer ANN models, each of which predicts one Fourier series coefficient of the cogging torque. Nevertheless, DistilBERT outperforms all of these models. This suggests that the use of natural language input plays a crucial role in the prediction task. Natural language descriptions provide richer contextual and background information than the purely numerical features used by the ANN models.Moreover, the textual input is more directly aligned with the target prediction task than the graph-based representations, which likely contributes to the superior performance of the BERT-based models.

TABLE VI. Training and inference efficiency comparison on a single NVIDIA A40 GPU. Training time is measured per epoch on 13653 samples, while inference time is measured on the full test set on 3875 samples.

Model	Train/Epoch	Inference
BERT-base-cased	∼74s	∼12s
DistilBERT	∼41s	∼7s
ResNet152	∼60s	∼10s
ANN 4-layer	∼ 1s	∼ 0.3s

C. Joint Training

In addition, we explore a scenario that traditional ANN models struggle to handle: designs with varying numbers of parameters. ANN models require a fixed input size, making it difficult to process multiple motor designs simultaneously when those designs differ in parameter count.

In contrast, our proposed natural language description supports variable-length inputs, enabling it to seamlessly represent and handle motor designs with differing topologies. To evaluate this capability, we create a new dataset by combining the motor datasets of Flat and V-shape magnets. We then evaluate the performance of our proposed input format using two BERT-based models: DistilBERT and BERT. The experimental setup follows the configuration described in Section IV-A, and the results are presented in Table V. Both DistilBERT and BERT models demonstrate strong performance on the mixed dataset, confirming the flexibility and robustness of our approach in handling variable-length and multi-topology motor design descriptions, which are difficult to support using conventional ANN-based approaches with fixed input dimensions.

D. Efficiency Study

All the experiments were conducted on a single NVIDIA A40 GPU. Both BERT-base-cased and DistilBERT can be trained within a reasonable amount of time on a single GPU. As shown in Table VI, DistilBERT demonstrates higher computational efficiency, reducing both training and inference time while still maintaining strong predictive performance. In particular, DistilBERT reduces the per-epoch training time from approximately 74s to 41s and decreases inference time on the test set from approximately 12s to 7s compared to BERT-base-cased, which is also comparable with ResNet152.

Since the ANN-based models require extensive hyperparameter tuning, we additionally evaluate their optimization cost using hyperparameter search. For the 4-layer ANN model, we tune the network width, batch size, learning rate, and number of training epochs. With a typical setting of 100 Optuna trials and pruning enabled, the total hyperparameter search takes approximately 3 hours, which is comparable to the training time of DistilBERT. In contrast, DistilBERT achieves strong performance without relying on extensive hyperparameter tuning.

V. Conclusion

In this paper, we proposed a motor surrogate modeling technique based on Text Transformer models. We first describe each motor design with natural language based on key design parameters, then fine-tune a BERT-based surrogate

model to predict the motor performances, cogging torque in particular. We have demonstrated with numerical tests that BERT models, combined with our proposed input format, significantly outperform variations of ANN models in cogging torque prediction on two datasets of interior permanent magnet synchronous motors. The flexibility and adaptability of the model is also demonstrated by the joint training on the two types of IPMSMs. The proposed framework can be extended to include different variations of motor topologies and trained to predict other motor performance metrics.

REFERENCES

[1] R. C. P. Silva, T. Rahman, M. H. Mohammadi, and D. A. Lowther, "Multiple operating points based optimization: Application to fractional slot concentrated winding electric motors," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1719–1727, 2017.

[2] S. Doi, H. Sasaki, and H. Igarashi, "Multi-objective topology optimization of rotating machines using deep learning," *IEEE transactions on magnetics*, vol. 55, no. 6, pp. 1–5, 2019.

[3] A. Khan, V. Ghorbanian, and D. Lowther, "Deep learning for magnetic field estimation," *IEEE Transactions on Magnetics*, vol. 55, no. 6, pp. 1–4, 2019.

[4] W. Kirchgässner, O. Wallscheid, and J. Böcker, "Estimating electric motor temperatures with deep residual machine learning," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7480–7488, 2020.

[5] Y.-m. You, "Multi-objective optimal design of permanent magnet synchronous motor for electric vehicle based on deep learning," *Applied Sciences*, vol. 10, no. 2, p. 482, 2020.

[6] Y. Shimizu, S. Morimoto, M. Sanada, and Y. Inoue, "Automatic design system with generative adversarial network and convolutional neural network for optimization design of interior permanent magnet synchronous motor," *IEEE Transactions on Energy Conversion*, vol. 38, no. 1, pp. 724–734, 2022.

[7] A. Choudhary, D. Goyal, and S. S. Letha, "Infrared thermography-based fault diagnosis of induction motor bearings using machine learning," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1727–1734, 2020.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2019.

[9] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," 2020.

[10] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting bert for end-to-end aspect-based sentiment analysis," 2019.

[11] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W. tau Yih, "Dense passage retrieval for open-domain question answering," 2020.

[12] R. Nogueira and K. Cho, "Passage re-ranking with BERT," 2020.

[13] O. Khattab and M. Zaharia, "ColBERT: Efficient and effective passage search via contextualized late interaction over BERT," 2020.

[14] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," 2020.

[15] Y. Sakamoto, B. Wang, T. Yamamoto, and Y. Nishimura, "Permanent magnet motor torque waveform prediction using learned gap flux," in *2024 IEEE 21st Biennial Conference on Electromagnetic Field Computation (CEFC)*, pp. 1–2, 2024.

[16] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016.

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2020.

BIOGRAPHIES

Siyuan Sun received the B.S. degree in electrical and computer engineering from University of Dallas at Texas, Dallas, Tx, USA, in 2018, and the M.S. degree electrical and computer engineering from Texas A&M University, College Station, TX, USA in 2020. In 2021, he joined Iowa State University, Ames, IA, USA as a Ph.D student in computer science department. His research interests are transfer learning and few-shot learning in machine learning field.

Ye Wang Ye Wang received the B.S. degree in electrical and computer engineering from Worcester Polytechnic Institute, Worcester, MA, USA, in 2005, and the M.S. and Ph.D. degrees in electrical and computer engineering from Boston University, Boston, MA, USA, in 2009 and 2011, respectively. In 2012, he joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, where he had also previously completed an internship in 2010. His research interests are information theory, machine learning, signal processing, communications, and data privacy/security.

Toshiaki Koike-Akino received the Ph.D. degree from Kyoto University, Kyoto, Japan, in 2005. During 2006-2010, he was a Postdoctoral Researcher at Harvard University, Cambridge, MA, USA, and he is currently Distinguished Research Scientist at MERL, Cambridge, MA, USA. He was the recipient of the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award. He is a Fellow of Optica (formerly OSA).

Tatsuya Yamamoto received the M.E. degree in Electrical and Electronic Engineering Course from Shinshu-university, Nagano, Japan, in 2016, and the Ph.D. degree in Department of Mathematics and System Development from the same university in 2019. Since 2019, he has been with Mitsubishi Electric Corporation, Japan, where he is currently a head researcher. His research interests include electric motor design, AI-based optimization and machine learning.

Yusuke Sakamoto received B. Sc in 2009 and M. Sc degree in 2011, both from the University of Tokyo, Japan. He also received Ph.D degree in Mechanical Engineering from Kyoto University, Japan, in 2024. Since 2011, he has been engaged in the research and development of electromagnetic applications, including charged particle accelerators, electric motors, and magnetic sensing.

Bingnan Wang received his B.S. from Fudan University in 2003, and Ph.D. from Iowa State University in 2009, both in Physics. He has been with Mitsubishi Electric Research Laboratories (MERL) since 2009, and is now a Lead Research Scientist. His current research focuses on electric machine analysis, design optimization, fault diagnosis with both physics modeling and machine learning techniques.