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.

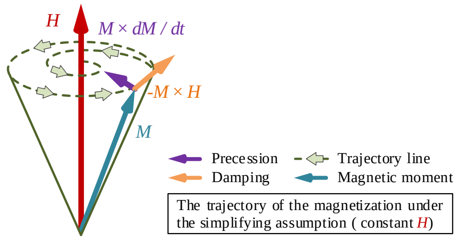

Fig. 1. The magnetization trajectory described by LLG equation.

The primary theoretical equation in micromagnetism is the Landau-Lifshitz-Gilbert (LLG) equation, which describes the motion of the magnetic moment, as expressed by:

$$\dot{m}(r,t) = -\gamma m(r,t)\times H_{eff} + \alpha m(r,t)\times \dot{m}(r,t) \tag{1}$$

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:

$$H_{eff} = H_{ext} + H_{exch} + H_{ani} + H_d \tag{2}$$

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 Ms. 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 Ms 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 ($B(t)\rightarrow H_{LLG}(t)$) 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.

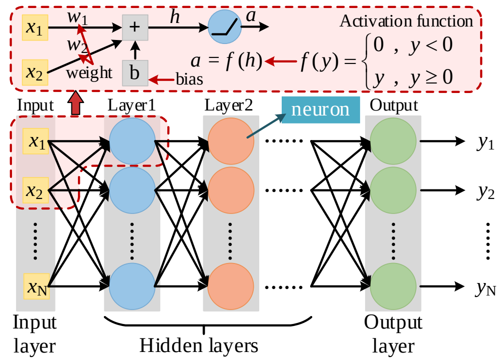

Fig. 2. A simple example of FCNN.

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.

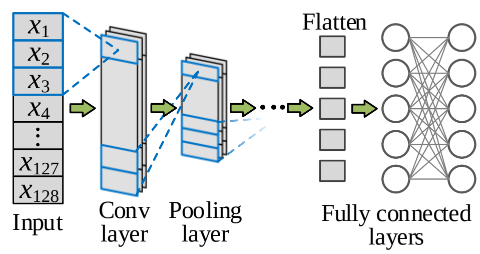

Fig. 3. A simple example of CNN.

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, Ms) into the learning process of the specially designed networks, the sub-model A can be regarded as the inverse model (B(t)→H_{LLG}(t)) of the LLG Equation with physical interpretability.

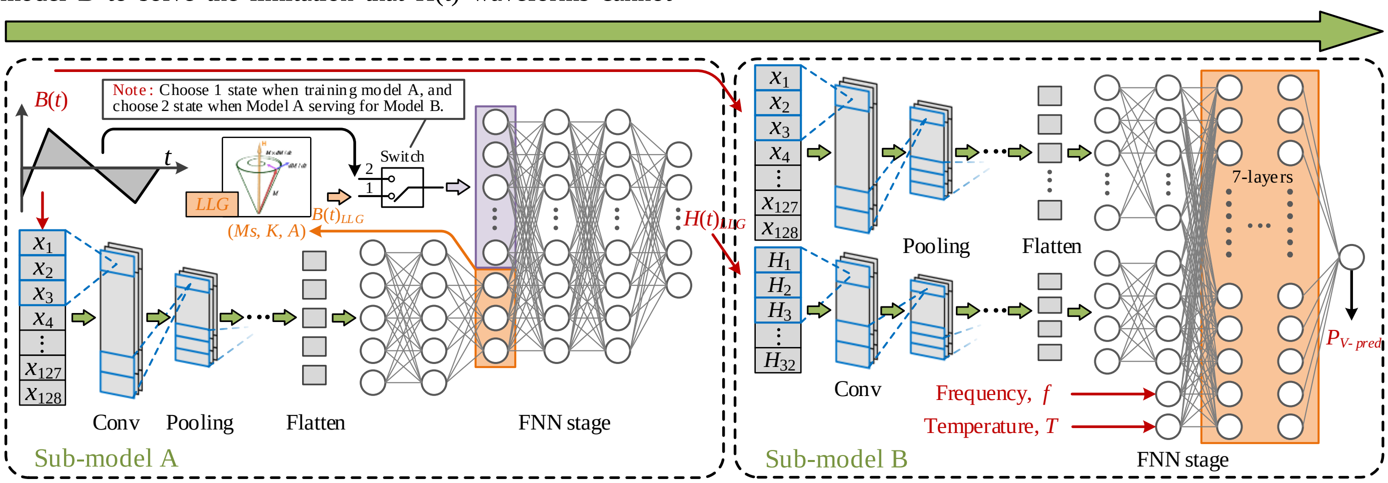

Fig. 4. The framework the proposed PI-MFF-CN method.

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)→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 (Ms, 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 (Ms, K, A). Before entering the next FCNN stage, MagTense simulates B_{LLG}(t) waveform based on the coefficients (Ms, 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 (Ms, 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 ℃ to 120 ℃, 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:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \tag{6}$$

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.

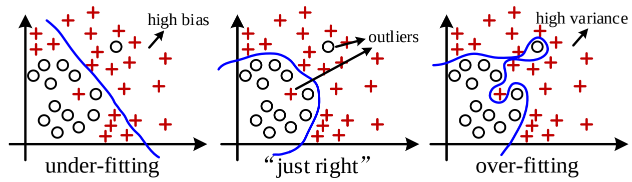

Fig. 5. Different fitting states of the model during training.

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:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\, y_i - \bar{y}_i \,| \tag{7}$$

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 *SmoothL₁* loss function, which is defined by:

$$SmoothL_1 = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y_i - \bar{y}_i)^2 & ,|\, y_i - \bar{y}_i \,| \le 1 \\ \frac{1}{n}\sum_{i=1}^{n}(|\, y_i - \bar{y}_i \,| - \frac{1}{2}) & ,|\, y_i - \bar{y}_i \,| > 1 \end{cases} \tag{8}$$

where δ is a hyperparameter. The *SmoothL₁* 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 SmoothL1, as follows:

$$J = \min J(\theta) = SmoothL_1 + \lambda \cdot \sum_{j=1}^{n}\theta_j^2 \tag{9}$$

where $\lambda \cdot \sum_{j=1}^{n}\theta_j^2$ 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. Reduce-LROnPlateau optimizer in Pytorh 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, Ms) into the gradient learning process of the network, the physics-inspired network can be regarded as the inverse model ($B(t)\rightarrow H_{LL}$G(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. Qinqin, "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.