xLSTM-Based Excitation Current Prediction for Synchronous Machines Towards Electro-spindle Drives

CHE Zhongyuan[1,2,3], *PENG Chong*[1,2,3*], *ZHANG Rui*[1], *WANG Chi*[1]

1. School of Mechanical Engineering and Automation, Beihang University, Beijing 102206, P. R. China; 2. Jiangxi Research Institute, Beihang University, Nanchang 330096, P. R. China; 3. Jiangxi Province Key Laboratory of High-End CNC Machine Tools, Jiangxi Research Institute, Beihang University, Nanchang 330096, P. R. China

(Received 6 July 2025; revised 20 January 2026; accepted 1 March 2026)

Abstract: Accurate excitation current prediction is crucial for the high-performance control of synchronous machines (SMs), which are widely employed in industrial drives such as electro-spindles. However, achieving accurate and generalizable prediction across multiple operating points is challenging due to coupled nonlinearities like thermal drift and magnetic saturation. This study proposes a novel prediction model based on the extended long short-term memory (xLSTM) network. The model integrates scalar LSTM (sLSTM) and matrix LSTM (mLSTM) units and leverages an exponential gating mechanism to enhance the capability for learning complex nonlinear mappings and long-term dependencies. Specifically, the vectorized parallel memory structure of sLSTM is suited to capturing slow parameter variations caused by thermal drift, while the matrix associative memory mechanism of mLSTM excels at learning multi-variable nonlinear coupling effects such as magnetic saturation. These two modules form a complementary hybrid architecture. Comparative analyses against traditional LSTM and gate recurrent unit (GRU) benchmarks were conducted using SM monitoring data covering various load and excitation conditions. In addition, an ablation study was performed using xLSTM with varying blending ratios of scalar and matrix LSTM components. Evaluation based on multiple error metrics and computational time demonstrates that the proposed xLSTM achieves superior accuracy, stronger generalization, lower computational overhead, and higher prediction stability. The underlying mechanisms are analyzed from architectural and algorithmic perspectives. These findings offer a novel data-driven modeling approach for SM excitation current, with potential value for applications requiring high-fidelity motor state estimation.

Key words: synchronous machine; excitation current modeling; data-driven method; extended long short-term memory (xLSTM)

CLC number: TM34 **Document code:** A **Article ID:** 1005-1120(2026)03-0443-20

0 Introduction

Synchronous machines (SMs) constitute a critical part of modern electromechanical drive systems, distinguished by their high-power density, precise speed synchronization, and outstanding torque characteristics[1-2]. These attributes make them the preferred choice for demanding applications, particularly in high-precision fields such as electro-spindle drives and renewable energy generation[3-4]. The performance of SMs in these diverse scenarios is fundamentally governed by the precise management of the rotor excitation current[5].

The excitation current directly controls the motor's magnetic field strength. Optimal excitation minimizes losses and maximizes efficiency, while deviations can lead to increased heating, reduced torque output, or instability[6]. Therefore, accurate prediction of the excitation current under varying operational conditions is a critical enabler for advanced

*Corresponding author, E-mail address: pch@buaa.edu.cn.

How to cite this article: CHE Zhongyuan, PENG Chong, ZHANG Rui, et al. xLSTM-based excitation current prediction for synchronous machines towards electro-spindle drives[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2026, 43(3):443-462.

http://dx.doi.org/10.16356/j.1005-1120.2026.03.009

control strategies, efficiency optimization, and condition monitoring.

However, constructing a robust prediction model is non-trivial. In practice, SMs operate across a wide range of load and speed setpoints. Even at steady-state operating points, the relationship between measurable variables and the excitation current is influenced by complex and coupled nonlinear phenomena. These include winding resistance variation due to thermal drift, magnetic saturation in the core, and the effects of manufacturing tolerances[7-8]. This multi-dimensional and non-stationary nature makes traditional model-based approaches challenging to apply effectively[9-10].

Traditional excitation current control methods exhibit significant limitations in electro-spindle applications[11]. Static equivalent circuit models are only suitable for analysis at a single steady-state point. They fail to learn the nonlinear mapping relationships across different operating conditions.

Model-based state estimation methods, such as the extended Kalman filter (EKF) and sliding mode observer (SMO), have been widely adopted for motor state estimation[12-13]. These methods rely on an accurate mathematical model. When the model parameters are precisely known and the system dynamics align with the model assumptions, they can provide robust estimates. However, obtaining and maintaining such a precise model is challenging due to coupled nonlinearities[8]. The performance of EKF can degrade due to model mismatches and the computational overhead associated with real-time Jacobian calculations[14]. Consequently, while effective in model-certain contexts, these model-dependent approaches face limitations in scenarios where the model is difficult to define or is subject to variation. These challenges highlight the need for alternative approaches that can circumvent the dependency on precise first-principles models.

Motivated by these limitations and the rapid advances in artifical intelligence (AI), data-driven methods have emerged as promising solutions to address complex nonlinear control challenges[15]. For high-dimensional nonlinear systems like SMs, AI methods demonstrate potential beyond traditional models. They excel at learning complex static nonlinear mappings, modeling implicit dependencies between operating conditions, and achieving cross-condition generalization[16-17].

In the domain of electrical machine modeling, Zeng et al.[18] proposed an integrated approach for six-phase hybrid-excited synchronous machines (6P-HESMs), combining finite element analysis (FEA) with data-driven experimental methods. This method incorporates vector space decomposition and incremental connection considerations to enable accurate torque estimation without direct phase current measurements. Similarly, focusing on compensation strategies, Lu et al.[19] developed a data-driven joint compensation technique that separately modeled magnetic saturation and core loss effects using nonlinear polynomials and neural networks. Model mismatch was calculated through differential analysis of low- and high-speed data, enhancing voltage and torque prediction accuracy at high speeds while establishing an efficient compensation framework for motor control.

Expanding structural analysis, Ciceo et al.[20] introduced a data-driven modeling method that updated material parameters via experimental modal analysis and synthesized frequency response functions (FRFs) using vibration synthesis techniques. A unified model for in-plane/out-of-plane modal behaviors and local effects was achieved by constructing a lookup table (LUT) through pole-zero representation. For thermal management applications, Sun et al.[21] presented a real-time estimation method for end-winding temperature in water-cooled interior permanent magnet synchronous motors (IPMSMs). The heat transfer problem with parameter variations was transformed into a linear time-invariant (LTI) system solved via Duhamel's theorem, where computational fluid dynamics (CFD) simulations derived response functions.

Addressing fault tolerance, Zhang et al.[22] developed an embedded control method that redistributed vector duty cycles in space vector pulse-width modulation (SVPWM). A voltage-vector-based current prediction model was established alongside transition-enabling mechanisms, reducing motor speed fluctuations and torque loss while improving post-fault performance. In fault diagnostics research, Guillén et al.[23] proposed a support vector regression (SVR)-based method to assess inter-turn fault severity in SM rotor windings. Features including stator voltage, active power, and reactive power were analyzed to predict theoretical excitation current, significantly improving fault detection sensitivity and accuracy over conventional models.

Concerning parameter estimation, Brosch et al.[6] created a torque and inductance identification algorithm using finite control set model predictive current control (FCS-MPCC). Recursive least squares were integrated with a hybrid motor model, leveraging FCS-MPCC's inherent system excitation for online parameter identification, differential inductance extraction, flux linkage determination, and torque computation. In comparative methodologies, Glŭčina et al.[24] evaluated current estimation performance across data-driven approaches. Multiple machine learning algorithms were tested using public datasets spanning diverse motor operating conditions, and results identified XGBoost as optimal for excitation current estimation accuracy and stability.

Despite existing advances, there remains room for improvement in balancing model complexity, computational efficiency, and cross-operating-condition generalization accuracy for the continuous multi-steady-state conditions faced by motorized spindles. For instance, finite element-based methods offer high accuracy but incur substantial computational costs, while many data-driven models are trained on single or limited operating conditions[25]. Thermal conduction equation methods depend on restrictive assumptions, limiting their applicability. Model predictive control techniques exhibit sensitivity to iron losses, impairing estimation accuracy during standstill or low-speed operation.

Thus, further research is required to develop novel data-driven methods for the modeling and temperature prediction of SMs. Such methods are expected to feature low computational overhead, high accuracy and few assumptions. The critical first step is to conduct comparative evaluation of the developed predictive models and realize high-accuracy, generalizable prediction of the field current for SMs under variable load conditions.

This study proposes a prediction model based on extended long short-term memory (xLSTM) for excitation current in SMs. A hybrid xLSTM architecture was developed by integrating scalar LSTM and matrix LSTM components at a 1∶1 ratio to form an xLSTM block. The network structure comprised 64 such blocks to constitute an xLSTM layer. The proposed xLSTM theoretically enhances the model's capability to handle long-term dependencies and complex nonlinear patterns. It establishes a mapping from multidimensional input features to the excitation current. For comparative analysis, benchmark models using LSTM and gate recurrent unit (GRU) architectures were implemented. Ablation studies were also conducted using xLSTM with varying blending ratios of scalar and matrix LSTM components to analyze their respective contributions. The experimental data were sourced from synchronous machine monitoring samples in the Kaggle database, covering operational data under various load and excitation current combinations. Prediction experiments evaluated model performance using multiple error metrics: Root mean squared error (RMSE), root mean squared error (MAE), mean bias error (MBE), R^2, and adjusted R^2, alongside computation time. The model's predictive accuracy and generalization performance, as well as the influence of cell architecture on learning efficacy and computational efficiency, are examined systematically.

The remainder of this paper is organized as follows. Section 1 introduces the fundamentals of xLSTM. It details the characteristics and implementations of both the scalar LSTM and the matrix LSTM. The design rationale for their blending ratios is explained, along with the motivation for building the excitation current prediction model. Section 2 describes the data sources and the experimental setup for data acquisition. It also defines the error and temporal metrics used to evaluate predictive performance. Section 3 presents the parameters and architectures of the comparative and ablation models.

The predictive characteristics of the proposed method are analyzed from structural, algorithmic, and training perspectives, discussing the underlying reasons. Finally, the paper concludes with a summary of the contributions and findings, and suggests directions for future research.

1 Architecture and Mechanisms of xLSTM

LSTM networks, introduced in the late 1990s, addressed a critical challenge in recurrent neural networks (RNNs) by mitigating the vanishing gradient problem through gated mechanisms that regulated information flow. The schematic diagram of the LSTM network structure is shown in Fig.1, where t denotes the current time step, $\boldsymbol{x}_t$ the input vector, and $\boldsymbol{h}_t$ the hidden state. The hyperbolic tangent activation function is denoted by tanh, while σ represents the sigmoid function. "+" denotes the element-wise addition and "×" the element-wise multiplication. Despite their foundational significance, traditional LSTMs exhibit inherent limitations: (1) Scalability bottlenecks in parallel computation due to sequential state updates, (2) representational constraints of scalar memory cells, and (3) inefficiencies in modeling complex dependencies over extended sequences.

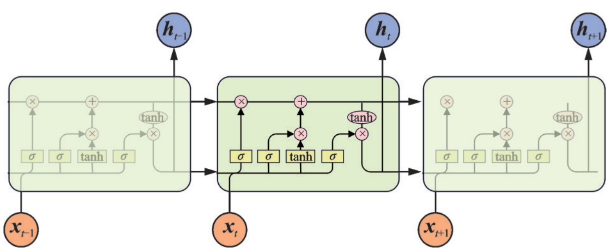

Fig.1 Schematic diagram of the LSTM network structure

To overcome these limitations, the xLSTM[26] architecture was developed through the fundamental reimagining of core LSTM components, with concurrent improvements in performance, scalability, and expressiveness. This section examines the principles underpinning xLSTM, with emphasis on two pivotal innovations: (1) Exponential gating and (2) novel memory structures.

xLSTM retains the strengths of LSTMs. It features notably stable gradient propagation and long-range dependency learning, while substantially improving capacity, parallelizability, and modeling efficacy for contemporary large-scale sequence processing tasks.

1.1 Replacing sigmoid with exponential gating

The gating mechanism constitutes the core operational component of any LSTM, governing three critical functions: (1) The integration rate of new information into the memory state (input gate), (2) the retention rate of prior information (forget gate), and (3) the information selection for output (output gate). Traditional LSTMs employ sigmoid activation functions (σ) for these gates, restricting outputs to the range (0, 1).

Sigmoid activations exhibit saturation behavior. When input magnitudes become substantially large (positive or negative), the sigmoid derivative approaches zero, inducing vanishing gradients during training. It impedes model learning, particularly for long-range temporal dependencies or abrupt state transitions. The bounded (0,1) output range inherently constrains gate influence.

xLSTM substitutes sigmoid activations in input and forget gates with exponential functions (exp). This fundamental modification yields significant advantages.

The exponential function's derivative remains self-replicating. For large positive gate inputs, the derivative magnitude persists, substantially reducing the vanishing gradient effect associated with sigmoid saturation. It facilitates enhanced learning efficacy across extended sequences.

Exponential gates generate outputs within (0, ∞). This permits dynamically scaled gate influence, where substantial positive inputs produced high-magnitude gate values. Consequently, decisive memory state updates become achievable, enabling either rapid information integration (sharp memorization) or complete state reset.

Large positive gate values collectively accelerate state transitions and expand the memory cell's dynamic range. This provides superior parametric control over both update velocity and amplitude within the memory subsystem.

Below is the mathematical formula of the input gate mechanism of xLSTM (compared with the standard LSTM).

Traditional LSTM is defined as

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{1}$$

xLSTM is defined as

$$\boldsymbol{i}_t = \exp(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{2}$$

where $\boldsymbol{W}_i$ are the weights and $\boldsymbol{b}_i$ the biases; $\boldsymbol{h}_{t-1}$ is the previous hidden state, $\boldsymbol{x}_t$ the current input, "$\cdot$" the matrix multiplication, σ the sigmoid function, and exp the exponential function.

Exponential gating reconfigures the memory cell from an information-preservation mechanism (via bounded gates) to a system capable of high-confidence rapid integration of significant new information when input conditions justify such action.

1.2 Innovations including sLSTM and mLSTM

The scalar memory cell in traditional LSTMs limits the model's capacity to represent complex, multi-faceted state information. xLSTM introduces two distinct novel memory architectures using exponential gating, operating in complementary ways.

The scalar LSTM (sLSTM) variant maximizes parallelizability during training and inference while mitigating the potentially large values generated by exponential gating. sLSTM maintains a single vector memory state ($\boldsymbol{c}_t$), analogous to a vectorized arrangement of traditional LSTM cells. Crucially, it decouples the memory update mechanism for each vector element, enabling parallel computation across the entire sequence dimension.

To ensure numerical stability during learning, normalization is integrated directly into the state update equations of sLSTM. This approach counters the large-value instability inherent to exponential gates, with layer normalization (LayerNorm) applied before gating activation being the most prevalent and effective solution. Below are the mathematical representations.

Gate input with pre-normalization is defined as

$$\boldsymbol{g}_t = \mathrm{LayerNorm}(\boldsymbol{W}_g \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_g) \tag{3}$$

Input gate is defined as

$$\boldsymbol{i}_t = \exp(\boldsymbol{g}_t) \tag{4}$$

Forget gate is defined as

$$\boldsymbol{f}_t = \exp(\boldsymbol{g}_t) \tag{5}$$

New candidate value is defined as

$$\boldsymbol{z}_t = \boldsymbol{W}_z \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_z \tag{6}$$

Memory state update is defined as

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{z}_t \tag{7}$$

Hidden state output is defined as

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \boldsymbol{c}_t \tag{8}$$

where $\boldsymbol{W}_g$ and $\boldsymbol{W}_z$ are the weight matrices and $\boldsymbol{b}_g$ and $\boldsymbol{b}_z$ the bias vectors; LayerNorm is the layer normalization function, $\odot$ the element-wise multiplication, $\boldsymbol{c}_{t-1}$ the previous memory state, and $\boldsymbol{o}_t$ the output gate.

The normalization and subsequent gating and update operations for each element in the memory vector $\boldsymbol{c}_t$ operate independently, conditioned on $\boldsymbol{h}_{t-1}$ and $\boldsymbol{x}_t$. Consequently, the entire memory state update at each time step can be computed in parallel across the sequence dimension. It represents a significant departure from the inherently sequential updates of traditional LSTMs and enables enhanced training efficiency on modern hardware architectures.

Regarding characteristics, the sLSTM variant excels in tasks demanding computational efficiency and robust parallelization. While normalization stabilizes learning dynamics, it introduces additional computational overhead. Crucially, memory elements remain decoupled at the element-wise level.

In contrast, the matrix LSTM (mLSTM) variant employs a fundamentally distinct memory structure to achieve substantially higher capacity and enable efficient associative recall. This approach transcends the limitations of scalar or vector memories by replacing $\boldsymbol{c}_t$ with a matrix memory state $\boldsymbol{C}_t$ of dimensions $d \times d$, where d denotes the model dimension. The matrix functions as an associative memory store.

The core innovation of mLSTM resides in its update mechanism, which utilizes an outer product-based covariance update rule. The normalized inputs for keys and values adopt the same form as Eq.(3), and the remaining variables are computed as follows.

Key vector is defined as

$$\boldsymbol{k}_t = \boldsymbol{W}_{\mathrm{k}} \cdot \boldsymbol{g}_t \tag{9}$$

Value vector is defined as

$$\boldsymbol{v}_t = \boldsymbol{W}_{\mathrm{v}} \cdot \boldsymbol{g}_t \tag{10}$$

Input gate, scalar is defined as

$$\boldsymbol{i}_t = \exp(\boldsymbol{W}_{\mathrm{i}} \cdot \boldsymbol{g}_t + \boldsymbol{b}_{\mathrm{i}}) \tag{11}$$

Forget gate, scalar is defined as

$$\boldsymbol{f}_t = \exp(\boldsymbol{W}_{\mathrm{f}} \cdot \boldsymbol{g}_t + \boldsymbol{b}_{\mathrm{f}}) \tag{12}$$

Memory state update is defined as

$$\boldsymbol{C}_t = \boldsymbol{f}_t \cdot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \cdot (\boldsymbol{k}_t \otimes \boldsymbol{v}_t) \tag{13}$$

Hidden state output is defined as

$$\boldsymbol{h}_t = \boldsymbol{C}_t \cdot \boldsymbol{q}_t \tag{14}$$

Query vector is defined as

$$\boldsymbol{q}_t = \boldsymbol{W}_{\mathrm{q}} \cdot \boldsymbol{g}_t \tag{15}$$

where $\boldsymbol{W}_{\mathrm{q}}$, $\boldsymbol{W}_{\mathrm{k}}$, and $\boldsymbol{W}_{\mathrm{v}}$ denote the weight matrices for the query, key, and value projections, respectively; $\boldsymbol{W}_{\mathrm{f}}$ and $\boldsymbol{W}_{\mathrm{i}}$ the weight matrices for the forget gate and input gate; and $\boldsymbol{b}_{\mathrm{f}}$ and $\boldsymbol{b}_{\mathrm{i}}$ the corresponding bias vectors. And, outer product ($\boldsymbol{k}_t \otimes \boldsymbol{v}_t$) generates a rank-1 matrix of dimensions $d \times d$, and each element (i, j) corresponds to the product $k_{ti} \times v_{tj}$. The matrix models the feature-level association between the key vector $\boldsymbol{k}_t$ and value vector $\boldsymbol{v}_t$ at time t. The matrix memory $\boldsymbol{C}_t$ is updated via a weighted summation: the prior state $\boldsymbol{C}_{t-1}$ scaled by the scalar forget gate $\boldsymbol{f}_t$, combined with the new association matrix ($\boldsymbol{k}_t \otimes \boldsymbol{v}_t$) scaled by the scalar input gate $\boldsymbol{i}_t$. This mechanism enables cumulative association storage within a unified high-capacity matrix. The hidden state $\boldsymbol{h}_t$ is computed as $\boldsymbol{C}_t \cdot \boldsymbol{q}_t$, where $\boldsymbol{q}_t$ denotes a query vector. This operation performs content-based addressing by retrieving associations from $\boldsymbol{C}_t$. The resultant vector $\boldsymbol{h}_t$ represents a weighted summation of values v, with weights proportional to the similarity between corresponding keys k and the query q.

Relative to sLSTM/traditional LSTM architectures, mLSTM exhibits significantly greater memory capacity ($O(d^2)$ versus $O(d)$) and demonstrates strong performance in tasks requiring complex pattern matching, relational reasoning, or content-based information retrieval. While the outer product update incurs higher computational cost than sLSTM's element-wise approach, it facilitates unique capabilities.

1.3 Integration and block design

xLSTM represents an integration of sLSTM and mLSTM rather than a simple summation. The hierarchical relationships and component structures of xLSTM, sLSTM, and mLSTM are visually demonstrated in Fig.2: (1) The original LSTM memory cell features a constant error carousel and gating mechanisms; (2) sLSTM and mLSTM memory cells introduce exponential gating, where sLSTM provides a new memory mixing technique and mLSTM enables full parallelization through a matrix memory cell state with a covariance-based update rule; (3) integration of mLSTM and sLSTM into residual blocks forms xLSTM blocks; (4) xLSTM blocks are stacked to construct the xLSTM architecture.

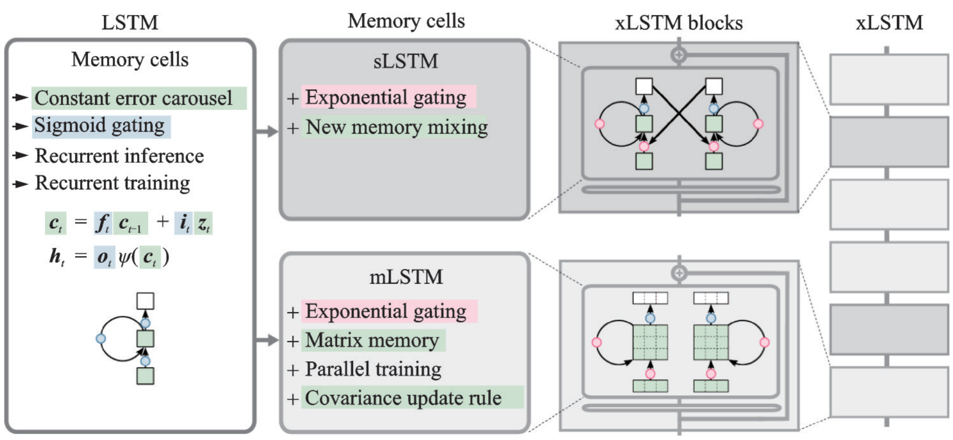

Fig.2 Structural diagrams and interrelationships of xLSTM, sLSTM, and mLSTM

These variants are architecturally unified into cohesive components within larger architectures through the following mechanisms.

(1) Residual framework integration: xLSTM blocks are conventionally embedded within residual structures. Given input $\boldsymbol{x}$, the output is computed as $\boldsymbol{x}+\boldsymbol{h}_t$, where $\boldsymbol{h}_t$ denotes the sLSTM/mLSTM output. This residual design facilitates deep network training.

(2) Projection layers: Inputs ($\boldsymbol{x}_t$) and optionally recurrent states ($\boldsymbol{h}_{t-1}$) are projected into higher-dimensional spaces prior to xLSTM gating and memory operations. Outputs ($\boldsymbol{h}_t$) may subsequently undergo down-projection to target dimensions.

(3) Gating and normalization: Layer normalization preconditions exponential gating for numerical stability. The output gate typically employs sigmoid or tanh activations.

(4) Variant interleaving: Practical xLSTM implementations frequently alternate sLSTM and mLSTM blocks. sLSTM specializes in sequence processing efficiency, while mLSTM delivers high-capacity associative memory. Block ratio and placement constitute task-specific hyperparameters.

(5) Vertical stacking: Model depth and representational capacity are enhanced by vertically stacking multiple xLSTM blocks.

For this study, the xLSTM architecture was implemented using one sLSTM and one mLSTM block at a 1∶1 integration ratio.

1.4 Analysis of the sLSTM and mLSTM blending ratio

In xLSTM, the blending ratio between the sLSTM and mLSTM is a critical architectural hyperparameter. It influences both the model's representational capacity and its computational efficiency. The sLSTM excels in highly parallelizable sequence processing and offers stable gradient propagation. This makes it suitable for capturing dependencies within the excitation current dynamics. Conversely, the mLSTM achieves a large-capacity associative memory through its matrix-based state, enabling the modeling of complex nonlinear interactions.

Therefore, a balanced 1∶1 ratio prevents either component from dominating the architecture. The resulting model leverages the sLSTM to learn temporal coherence in variations while utilizing the mLSTM to establish feature-level associations.

Furthermore, the residual framework within each xLSTM module facilitates gradient flow in the deep architecture. The balanced composition helps maintain consistent gradient magnitudes across both the sLSTM and mLSTM pathways. This mitigates the risks of vanishing or exploding gradients during training.

Beyond the above analysis, Section 3.3 details ablation experiments. These experiments compare the performance of the balanced ratio against asymmetric configurations (sLSTM-to-mLSTM ratios of 2∶1, 1∶2, 4∶1, and 1∶4). The comparison focuses on their error metrics, generalization capability, computational time, and distinctive characteristics in modeling excitation current.

1.5 Comparative explanation of xLSTM

In contrast to traditional LSTMs, xLSTM provides fundamental enhancements through exponential gates, which both more effectively resolve vanishing gradient issues and significantly enhance memory structures. Specifically, sLSTM employs vector-based parallel memory, while mLSTM utilizes associative matrix memory, both representing substantial advances over traditional scalar cells.

Compared to Transformers which exclusively rely on self-attention for sequence modeling and thus enable full parallelization but lack recurrent states, xLSTM offers distinct advantages, such as enhanced memory retention. Meanwhile, the sLSTM variant achieves $O(n)$ parallel training complexity, and mLSTM attains $O(nd^2)$, outperforming the $O(n^2d)$ complexity of standard Transformer attention, where n denotes the sequence length.

Furthermore, xLSTM maintains a recurrent state and demonstrates superior performance on very long sequences and associative recall benchmarks. Notably, mLSTM's associative recall mechanism provides a potentially more efficient content-based addressing approach than dot-product attention.

Relative to other modern RNNs, xLSTM exhibits architectural and performance distinctions.

For example, state space models (SSMs) employ linear time-invariant systems and convolution, fundamentally differing from xLSTM's gated non-linear recurrence and matrix memory. Similarly, the time-mixing and channel-mixing operations of receptance weighted key value (RWKV) differ from xLSTM's exponential gating framework and hybrid sLSTM/mLSTM memory architecture in their approach to sequence modeling and memory retention.

1.6 Physical motivation for applying xLSTM to excitation current prediction

Applying xLSTM to model the excitation current of SMs is not a simple model substitution. It is motivated by a correspondence between the mathematical properties of its core components and the physical characteristics of excitation current dynamics.

During operation, a synchronous machine's excitation current is a complex variable determined by multi-physics coupling. The temperature rise caused by power losses, with time constants ranging from minutes to hours, leads to slow drifts in winding resistance and magnetic circuit saturation. These drifts affect electrical parameters and primarily govern the trend and slowly varying components of the excitation current. Conversely, load fluctuations, faults, or changes in control commands can induce electromagnetic transients. It demands that the model accurately learn the nonlinear coupling relationships between the excitation current and other measurable variables.

The hybrid architecture of xLSTM provides a well-matched and balanced modeling framework for this purpose. The structure of sLSTM is well-suited to model the slow thermal dynamics from power losses, capturing the gradual variations in excitation current. Its scalar memory cell and exponential gating grant it long-term dependency retention. The exponential gating controls the rate at which historical information is retained or forgotten.

In parallel, the matrix memory of mLSTM forms a high-dimensional associative storage. Its covariance update rule enables the model to efficiently learn and store the complex mappings between excitation current and multiple variables across different operating states in a parallelizable manner. It facilitates agile responses to dynamic changes. Consequently, the dual-module design of xLSTM structurally aligns well with the requirements for modeling excitation current.

2 Experimental Data and Error Metrics

2.1 Data description

Synchronous machines can operate across a wide range by regulating excitation current, achieving leading, lagging, or unity power factor. The flexibility enhances the motor's own electrical efficiency. This study employs experimental data comprising SM monitoring samples under various load and excitation current combinations. Data acquisition commenced once the motor reached synchronous speed and was maintained under steady load. Such data are suitable for validating a model's capability in capturing fundamental nonlinear mappings and long-term dependencies. The xLSTM design incorporates its theoretical suitability for non-stationary operating conditions (refer to Section 1.6), which is further examined in later discussions (refer to Section 3.3).

Fig.3 provides a schematic overview of the experimental setup[27-28]. The device under test is a three-phase synchronous machine (M 3～), driven by an auxiliary induction motor to near-synchronous speed. The stator windings, accessed via terminals A, B, C, and neutral 0, are energized by a power-frequency AC supply and loaded by an adjustable inductive load. Stator line voltage V_L, and power factor $\cos\varphi$ are synchronously measured by the ammeter A, voltmeter V, and power factor meter, respectively. In the rotor excitation circuit, a variable resistor is connected in series to enable manual adjustment of the DC field current I_f, which is monitored by the ammeter A in the rotor loop. The experimental data are sourced from a real-world SM dataset (accessible at https://www.kaggle.com/datasets/fedesoriano/synchronous-machine-dataset, accessed on 9 April 2024). The dynamic operational data, collected by varying load and excitation settings, form the foundational dataset for the subse-

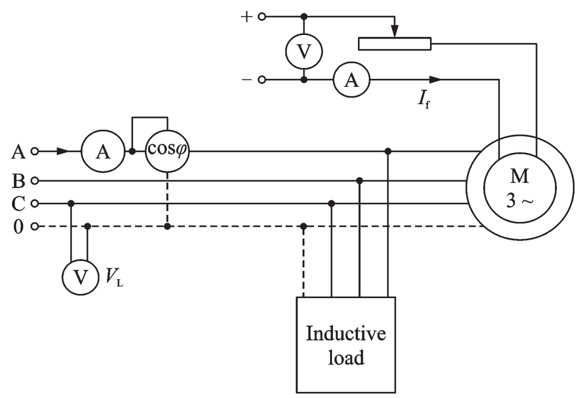

Fig.3 Schematic of the experimental workflow with a SM[27-28]

quent modeling work.

Specifically, the SM serves as the test object and the prime mover is an induction motor. The prime mover drives the SM to near-synchronous speed. A variable resistor, connected in series within the rotor excitation circuit, is used to manually adjust the DC excitation current. The load and power supply are an adjustable mechanical load and a power-frequency AC supply energizing the stator windings, respectively. Stator current, stator voltage, power factor, power factor error, and excitation current were synchronously measured and recorded using sensors and meters. Data acquisition was accomplished via a data acquisition card connected to an upper computer.

The experimental procedure for collecting operational data from the SM followed these steps.

(1) Drive and excitation preparation: In the test bench, an auxiliary motor was used to drive the SM under test into rotation. A rheostat was connected in series within the rotor excitation circuit to provide a continuously adjustable DC excitation supply.

(2) Stator power supply and startup: AC voltage was applied to the stator windings of the SM. The auxiliary motor drove it to a rotational speed very close to the synchronous speed.

(3) Excitation application and pull-in synchronization: DC voltage was applied to the rotor excitation windings. The interaction between the rotor magnetic field and the stator rotating magnetic field locked the rotor into synchronization, pulling the motor into synchronous operation.

(4) Establishing a baseline and conducting experiments: For each fixed load condition, the excitation current was adjusted via the rheostat to achieve a unity power factor operating point, where the stator current is minimized. This point served as the experimental baseline. Subsequently, with load and voltage held constant, the excitation current was increased (leading to over-excitation and leading power factor) or decreased (leading to under-excitation and lagging power factor) to collect data across a range of excitation levels.

The experimental procedure was repeated under multiple distinct load conditions. For each test, various input parameters and the output parameter of the SM were measured and recorded from the test bench. The resulting operational parameters, covering a wide range of excitation and load combinations, formed the dataset for model training and testing.

The dataset is publicly available on GitHub via the following link: https://github.com/567ZYC/Dataset-for-Excitation-Current-Prediction-in-Synchronous-Machines. Table 1 summarizes the physical meanings and interrelationships of the five key electrical parameters.

Table 1 Electrical parameters collected from SMs

Category	Physical variable	Description
Input	Load current	Representing the alternating current intensity in stator windings during motor operation, the reflecting torque demands from mechanical loads.
	Power factor	Describing the ratio of active power to apparent power, and indicating motor operational efficiency.
	Power factor error	Representing the deviation between measured power factor and target value, used for evaluating system control accuracy.
	Excitation current change	Representing the dynamic adjustment of excitation current recorded during regulation processes.
Target	Excitation current	Representing the direct current in rotor windings, and serving as the target variable for model prediction.

A SM is an AC motor characterized by its rotor rotating at a speed strictly synchronized with the supply frequency. Its operating principle relies on the interaction between the stator's rotating magnetic field and the rotor's magnetic field, which can be either generated by permanent magnets (in permanent magnet synchronous machines, PMSMs) or DC excitation (in electrically excited SMs).

The synchronous speed (n_s) in revolutions per minute is determined by the formula $n_s=(120\times f)/P$, where f is the supply frequency and P the number of poles. Equivalently, it can be expressed as $n_s=(60\times f)/p$, where p denotes the number of pole pairs ($p=P/2$). This constant relationship ensures that the rotor maintains a fixed speed proportional to the grid frequency.

The differences between SMs and another commonly used motor and the AC servo motor, are shown in Table 2. Therefore, inaccurate excitation current estimation can disrupt torque-field decoupling, leading to torque ripple and reduced machining quality. Precise estimation avoids over-excitation, minimizing stator copper losses and core eddy current losses. Closed-loop control of excitation current also contributes to reduced motor heating.

Table 2 Key differences between synchronous machines and AC servo motors

Category	Synchronous machine	AC servo motor
Operating principle	Rotor speed synchronizes with the stator's rotating magnetic field	Uses closed-loop control with feedback (encoder/resolver) for positioning/speed
Control method	Speed depends on supply frequency	Closed-loop control with algorithms
Efficiency	High efficiency at rated load and synchronous speed	High efficiency across variable speeds and loads
Feedback system	Often operates without feedback	Requires encoder/resolver for real-time position/speed feedback
Cost	Lower cost for simple designs; higher cost for large power ratings	Higher system cost
Maintenance	Robust design with minimal maintenance	Requires periodic maintenance for feedback devices and electronics
Power range	Wide range	Typically smaller
Torque characteristics	Constant torque at synchronous speed; limited overload capacity	High starting torque and overload capacity for rapid acceleration
Complexity	Simpler construction and control system	Complex control electronics and software integration

2.2 Error metrics

Multiple metrics are selected to evaluate both prediction accuracy and operational efficiency: RMSE quantifies discrepancies between predicted and true values; MAE computes average absolute deviations, demonstrating robustness to outliers; MBE analyzes systematic prediction bias, where positive/negative values indicate overall underestimation/overestimation; R^2 measures the proportion of target variance explained by the model; Adjusted R^2 corrects R^2 inflation with added features; Runtime is total time required for one cycle.

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{16}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{17}$$

$$\mathrm{MBE}=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i) \tag{18}$$

$$R^2=1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \tag{19}$$

$$\text{Adjusted } R^2=1-\left(\frac{(1-R^2)(N-1)}{N-k-1}\right) \tag{20}$$

$$\mathrm{Runtime}=\mathrm{Time}_{\mathrm{train}}+\mathrm{Time}_{\mathrm{test}} \tag{21}$$

where N denotes the total number of samples; y_i

and $\hat{y}_i$ represent the true value and predicted value of the sample, respectively; $\bar{y}$ is the arithmetic mean of all data samples; and k corresponds to the number of independent variables included in the model.

The integrated use of these metrics enables a balanced assessment of predictive accuracy. RMSE imposes a quadratic penalty on large errors, rendering it sensitive to outliers and severe prediction failures. MAE provides a linear, robust measure of average error magnitude that is resistant to outlier distortion, offering intuitive interpretability. MBE quantifies systematic bias. Values near zero indicate unbiased performance, while persistent deviations reveal fundamental model flaws.

Collectively, these metrics dissect error structure: RMSE captures volatility, MAE reflects typical accuracy, and MBE diagnoses directional trends. R^2 measures the proportion of variance explained by the model. Adjusted R^2 enhances this metric by penalizing superfluous features, mitigating overfitting from excessive complexity. Employing both metrics ensures robustness. R^2 evaluates absolute fit, while adjusted R^2 accounts for model parsimony. Models that achieve marginal accuracy gains at excessive runtime costs are often impractical in resource-constrained environments. Runtime directly quantifies computational burden.

Consequently, the integrated application of error and time metrics addresses the limitations of single-metric evaluation, delivering a comprehensive performance assessment through simultaneous examination of accuracy, robustness, and computational efficiency.

3 Experiments and Discussion

3.1 Baseline models and definitions

LSTM and GRU were selected as comparative baselines for xLSTM since they represent the most established recurrent network architectures for sequential and regression modeling[29-30]. This selection is further justified by their direct relevance to xLSTM's design objectives and potential improvement pathways.

Both LSTM and GRU address the long-term dependency limitations of traditional RNNs through gating mechanisms. Expanding on this foundation, xLSTM introduces an enhanced exponential gating mechanism. The approach directly mitigates the limitations of LSTM's sigmoid-based gating while retaining GRU's structural efficiency. Subsequently, the comparative evaluation assesses whether xLSTM achieves improvements in modeling accuracy and computational efficiency. Detailed architectures and parameters for all implemented models (xLSTM, LSTM, and GRU) are provided in Table 3.

Table 3 Comparison of network architectures and model configurations

Model	Network architecture	Hidden layer neurons	Learning rate	Learning rate decay	Epoch
xLSTM	Stacked layers (1 sLSTM, 1 mLSTM)	64	0.001	No decay	1 000
LSTM	Two stacked LSTM layers	64	0.001	No decay	1 000
GRU	Two stacked GRU layers	64	0.001	No decay	1 000

Model hyperparameters were primarily set according to well-established empirical values commonly adopted in similar modeling tasks. To ensure a fair comparison, the baseline models (LSTM, GRU) and the proposed xLSTM were designed with comparable parameter counts and network depths. This design choice prevents performance differences from being attributed to disparities in model scale. All models were trained using the Adam optimizer.

Although advanced optimization algorithms and extensive hyperparameter tuning could potentially improve performance, such approaches often require substantial computational resources and time. Since one key objective of this work is to develop an efficient and readily applicable prediction model, we prioritized architectural innovation over hyperparameter search.

Notably, even with these empirically chosen parameters, the proposed xLSTM achieved excellent predictive accuracy and computational efficiency. The outcome suggests that the xLSTM architecture possesses inherent suitability and robustness for the excitation current prediction task. Its performance stems primarily from its dual-module architecture (integrating sLSTM and mLSTM), not from extensive parameter tuning. Thus, the empirically configured model proved effective.

The model architectures of xLSTM, LSTM, and GRU are illustrated in Figs.4—6. From left to right, each diagram sequentially presents: Feature vector, input layer, hidden layer, fully connected layer (output layer), and prediction target. Consequently, these networks share a similar overall structure, differing primarily in their hidden layer configurations.

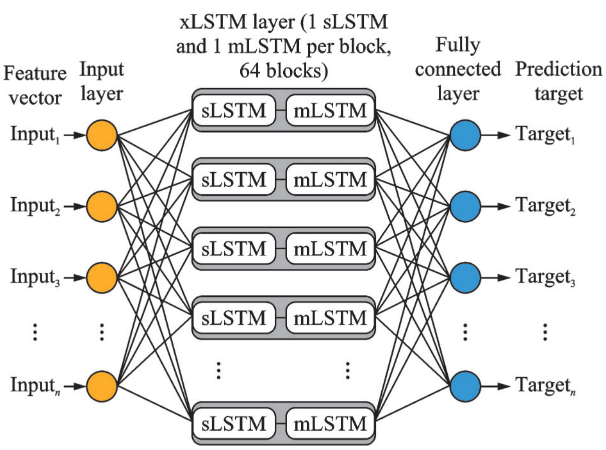

Fig.4 Architecture of the xLSTM model

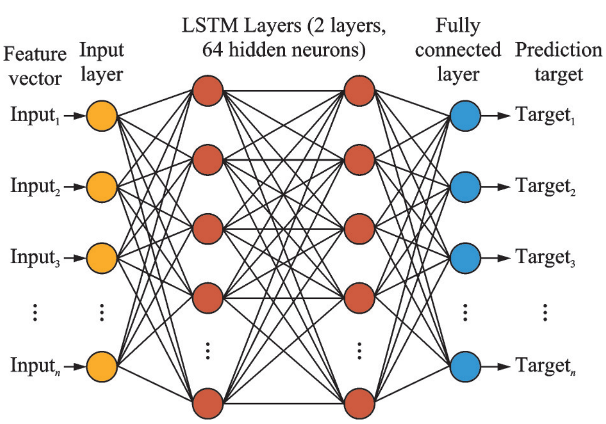

Fig.5 Architecture of the LSTM model

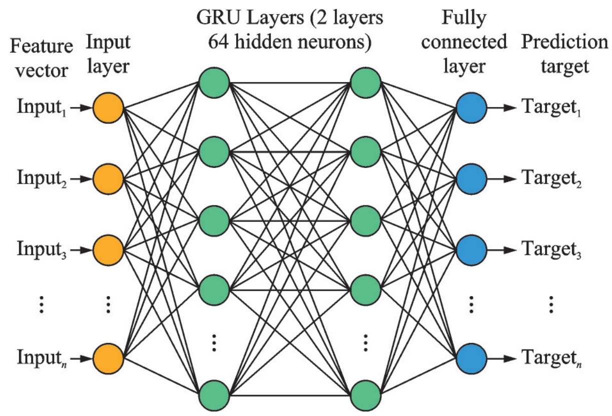

Fig.6 Architecture of the GRU model

Specifically, the three models employ distinct hidden layers: xLSTM layers, LSTM layers, and GRU layers, respectively. Each xLSTM cell integrates one sLSTM and one mLSTM module, resulting in a structure functionally analogous to a two-layer stacked network in traditional implementations. The LSTM and GRU layers consist of standard LSTM cells and GRU cells, respectively. All the architectures utilize 64 neurons in their hidden layers.

3.2 Prediction experiments and analysis of model characteristics

The three models were individually trained and tested using a 1∶1 training-to-test ratio. In each iteration, half of the samples were randomly selected for training while the remaining half comprised the test set. To mitigate randomness effects, twenty independent prediction experiments were conducted per model. Table 4 summarizes the mean, median, and standard deviation (STD) of the experimental results.

The xLSTM significantly outperformed both LSTM and GRU across all error metrics (RMSE, MAE, MBE, R^2, and Adjusted R^2) in both training and testing datasets. Notably, xLSTM achieved the shortest computation time (mean total time of 31.84 s), contradicting the conventional expectation that high accuracy requires substantial computational resources. These results demonstrate the architectural superiority of the combination of one sLSTM and one mLSTM over stacked LSTM or GRU ar-

Table 4 Descriptive statistics of experimental results

Model	Stage	Metric	Mean	Median	STD
xLSTM	Train	RMSE	0.000 4	0.000 4	0.000 1
		MAE	0.000 3	0.000 3	0.000 1
		MBE	−6.66E−06	1.45E−05	2.35E−04
		R^2	0.999 994	0.999 996	4.22E−06
		Adjusted R^2	0.999 994	0.999 996	4.28E−06
	Test	RMSE	0.000 4	0.000 4	0.000 2
		MAE	0.000 3	0.000 3	0.000 1
		MBE	−8.28E−06	3.98E−05	2.50E−04
		R^2	0.999 993	0.999 995	4.44E−06
		Adjusted R^2	0.999 993	0.999 995	4.51E−06
	Total	Time	31.835 5	31.747 2	0.291 9
LSTM	Train	RMSE	0.001 7	0.001 1	0.001 4
		MAE	0.001 5	0.001 0	0.001 1
		MBE	−4.18E−05	−3.51E−04	1.47E−03
		R^2	0.999 855	0.999 965	2.22E−04
		Adjusted R^2	0.999 853	0.999 964	2.26E−04
	Test	RMSE	0.001 7	0.001 1	0.001 3
		MAE	0.001 5	0.001 0	0.001 1
		MBE	−2.88E−05	−3.55E−04	1.45E−03
		R^2	0.999 854	0.999 960	2.21E−04
		Adjusted R^2	0.999 851	0.999 959	2.25E−04
	Total	Time	44.737 4	43.377 0	3.843 4
GRU	Train	RMSE	0.002 4	0.001 6	0.002 0
		MAE	0.002 1	0.001 4	0.001 7
		MBE	−4.27E−04	−2.93E−04	2.21E−03
		R^2	0.999 699	0.999 922	4.60E−04
		Adjusted R^2	0.999 695	0.999 921	4.67E−04
	Test	RMSE	0.002 4	0.001 6	0.002 0
		MAE	0.002 0	0.001 4	0.001 6
		MBE	−3.83E−04	−2.45E−04	2.15E−03
		R^2	0.999 708	0.999 918	4.58E−04
		Adjusted R^2	0.999 704	0.999 917	4.64E−04
	Total	Time	38.042 4	38.075 7	0.540 5

chitectures.

Furthermore, significant differences were observed in prediction stability. xLSTM exhibited error standard deviations generally 1—2 orders of magnitude lower than those of other models, indicating substantially more stable predictions. For SM excitation current modeling, enhanced stability translates to greater model reliability and increased practical value.

On the testing set, the xLSTM RMSE measured approximately 25% of the LSTM RMSE and 18% of the GRU RMSE. The R^2 value for xLSTM approached the theoretical maximum of 1.0 (xLSTM: 0.999 993 vs. LSTM: 0.999 854). Regarding systematic bias, the MBE metric revealed potential outliers and directional prediction instability. LSTM and GRU showed MBE values of -2.88×10^{-5} and -3.83×10^{-4}, respectively. Conversely, xLSTM demonstrated the smallest absolute deviation at -8.28×10^{-6}, outperforming other models by two orders of magnitude.

xLSTM demonstrates approximately 29% faster computational performance than LSTM and 16% faster than GRU. LSTM exhibits significantly higher time standard deviation (3.84 s) compared to other models, indicating reduced training stability. Its instability may originate from initialization sensitivity or gradient fluctuations. xLSTM achieves the lowest computational time standard deviation (0.291 9 s).

All three models employ a simple two-layer stacked architecture. These results indicate that for medium-to-large scale computations, xLSTM incurs substantially lower computational overhead than both LSTM and GRU. For excitation current estimation in SMs, computational efficiency and real-time inference are critical performance factors. The reduced computational overhead facilitates more responsive feedback.

Fig.7 visually compares computational efficiency and stability through a point-line diagram based on 20 experimental trials. xLSTM consistently achieves the shortest total training and testing duration across all trials. GRU exhibits longer execution times with significant fluctuations between trials 12 and 16. LSTM exceeds 50 s in its first three trials, with subsequent computations (trials 5—20) fluctuating between 40 and 45 s.

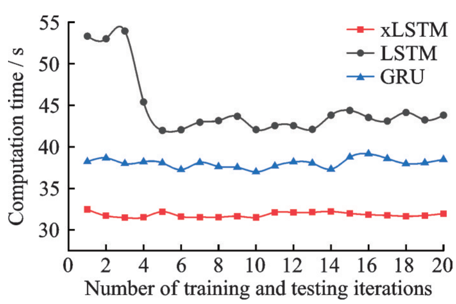

Fig.7 Line plot of computation time for xLSTM, LSTM, and GRU

Fig.8 displays a box plot with distribution curves, illustrating the MAE distribution across different models evaluated on the test set. The vertical span of each box reflects data concentration, aligning with the shape of the adjacent distribution curve.

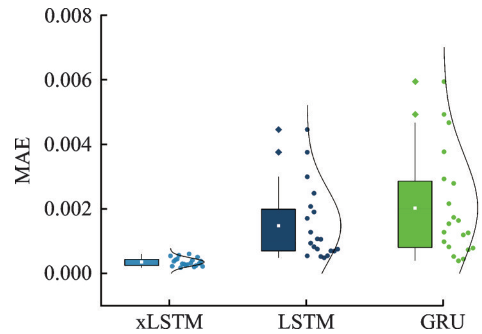

Fig.8 Boxplot of MAE on the test set overlayed with distribution curves

The white square within each box denotes the mean value, while the upper and lower whiskers represent the outlier range using a coefficient of 1. Circular markers indicate individual model MAE values on the test set, and diamond markers denote outliers.

xLSTM demonstrates the lowest mean MAE, exhibits no outliers, and displays the most concentrated MAE distribution. In contrast, GRU shows two diamond-shaped outliers beyond its upper whisker, indicating two instances of abnormally high MAE predictions across 20 experimental trials. Furthermore, GRU exhibits a highly dispersed MAE distribution, revealing inconsistent modeling accuracy for SM excitation currents.

Fig.9 displays a kite plot derived from the test set MBE values. The area of each kite corresponds to the absolute magnitude of the MBE, offering a visual representation of MBE performance and stability across the 20 experimental replicates. Table 4 lists the average MBE values for xLSTM, LSTM, and GRU as -8.28×10^{-6}, -2.88×10^{-5}, and -3.83×10^{-4}.

All three models exhibit an overall negative bi-

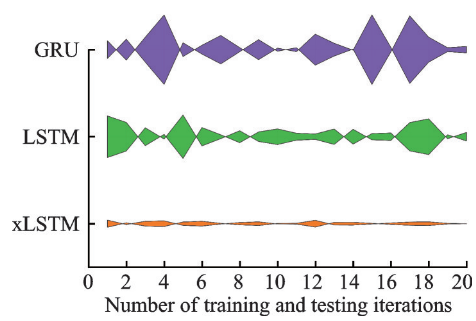

Fig.9 Kite plot of MBE for xLSTM, LSTM, and GRU on the test set

as. However, xLSTM demonstrates a significantly smaller MBE magnitude. Furthermore, Fig. 9 indicates that xLSTM maintains stable MBE values consistently near zero. On the contrary, substantial MBE magnitudes are observed for GRU in replicates 4, 15, and 17. While LSTM exhibits lower predictive accuracy and computational efficiency than GRU, its overall MBE remains smaller in magnitude. LSTM displays only minor MBE increases in replicates 1, 5, and 18.

The preceding analysis demonstrates that xLSTM achieves both the highest accuracy and computational efficiency. These findings are consistent with theoretical predictions. The architecture's novel exponential gating and memory mechanisms significantly improve parameter utilization efficiency and enhance excitation current modeling capabilities.

Notably, conventional LSTM exhibits the highest computational cost while achieving substantially lower accuracy than xLSTM. GRU delivers intermediate performance across these metrics. For practical implementations, xLSTM represents the preferred modeling approach, especially for real-time excitation current estimation in SMs due to its pronounced comprehensive advantages.

3.3 Ablation study on the sLSTM and mLSTM blending ratio

To further validate the rationale behind employing a 1∶1 blending ratio of sLSTM to mLSTM within the developed xLSTM, an ablation study was performed. We compared the following five architectural configurations.

(1) xLSTM-1∶1: Each module contains 1 sLSTM+1 mLSTM (the proposed scheme).

(2) xLSTM-2∶1: Each module contains 2 sLSTM+1 mLSTM.

(3) xLSTM-1∶2: Each module contains 1 sLSTM+2 mLSTM.

(4) xLSTM-4∶1: Each module contains 4 sLSTM+1 mLSTM.

(5) xLSTM-1∶4: Each module contains 1 sLSTM+4 mLSTM.

All configurations were implemented under the same residual module framework, with consistent overall network architecture (number of stacked modules, hidden units, training settings). Each configuration was trained and tested over 20 independent runs. The mean, median, and standard deviation of the experimental error and time metrics are presented in Table 5. The ablation study results reveal distinct model characteristics associated with different blending ratios. First, xLSTM (1∶1) achieved the best overall performance across all evaluated metrics: Prediction accuracy (RMSE, MAE, R^2), computational time, and stability (reflected by a low standard deviation).

Table 5 Ablation study results for xLSTM models with different blending ratios

Model	Stage	Metric	Mean	Median	STD
xLSTM (1∶1)	Train	RMSE	0.000 4	0.000 4	0.000 1
		MAE	0.000 3	0.000 3	0.000 1
		MBE	−6.66E−06	1.45E−05	2.35E−04
		R^2	0.999 994	0.999 996	4.22E−06
		Adjusted R^2	0.999 994	0.999 996	4.28E−06
	Test	RMSE	0.000 4	0.000 4	0.000 2
		MAE	0.000 3	0.000 3	0.000 1
		MBE	−8.28E−06	3.98E−05	2.50E−04
		R^2	0.999 993	0.999 995	4.44E−06
		Adjusted R^2	0.999 993	0.999 995	4.51E−06
	Total	Time	31.835 5	31.747 2	0.291 9

Continued

Model	Stage	Metric	Mean	Median	STD
xLSTM (2:1)	Train	RMSE	0.000 5	0.000 4	0.000 3
		MAE	0.000 4	0.000 3	0.000 2
		MBE	−7.03E−05	−4.21E−05	3.30E−04
		R^2	0.999 990	0.999 995	1.34E−05
		Adjusted R^2	0.999 990	0.999 995	1.36E−05
	Test	RMSE	0.000 5	0.000 4	0.000 3
		MAE	0.000 4	0.000 3	0.000 2
		MBE	−6.99E−05	−5.09E−05	3.32E−04
		R^2	0.999 990	0.999 994	1.45E−05
		Adjusted R^2	0.999 990	0.999 994	1.48E−05
	Total	Time	34.526 5	34.439 8	0.427 1
xLSTM (1:2)	Train	RMSE	0.000 6	0.000 6	0.000 3
		MAE	0.000 5	0.000 4	0.000 2
		MBE	1.83E−05	2.12E−05	4.12E−04
		R^2	0.999 985	0.999 990	1.39E−05
		Adjusted R^2	0.999 985	0.999 990	1.41E−05
	Test	RMSE	0.000 7	0.000 6	0.000 3
		MAE	0.000 5	0.000 5	0.000 2
		MBE	2.61E−05	1.59E−05	4.16E−04
		R^2	0.999 984	0.999 988	1.41E−05
		Adjusted R^2	0.999 984	0.999 988	1.43E−05
	Total	Time	34.176 7	34.051 3	0.825 2
xLSTM (4:1)	Train	RMSE	0.000 5	0.000 6	0.000 2
		MAE	0.000 4	0.000 4	0.000 2
		MBE	4.93E−05	8.24E−06	3.36E−04
		R^2	0.999 989	0.999 991	9.68E−06
		Adjusted R^2	0.999 989	0.999 991	9.82E−06
	Test	RMSE	0.000 6	0.000 6	0.000 2
		MAE	0.000 5	0.000 5	0.000 2
		MBE	6.32E−05	−2.01E−06	3.34E−04
		R^2	0.999 988	0.999 990	1.01E−05
		Adjusted R^2	0.999 988	0.999 990	1.02E−05
	Total	Time	34.287 1	34.126 9	0.473 4
xLSTM (1:4)	Train	RMSE	0.000 6	0.000 5	0.000 4
		MAE	0.000 5	0.000 4	0.000 4
		MBE	−4.25E−05	−1.06E−05	3.63E−04
		R^2	0.999 982	0.999 993	2.77E−05
		Adjusted R^2	0.999 982	0.999 993	2.81E−05
	Test	RMSE	0.000 6	0.000 5	0.000 4
		MAE	0.000 5	0.000 4	0.000 4
		MBE	−5.55E−05	−7.66E−05	3.57E−04
		R^2	0.999 982	0.999 992	2.72E−05
		Adjusted R^2	0.999 982	0.999 992	2.76E−05
	Total	Time	34.823 1	34.569 4	0.599 1

Performance gradually declined as the ratio deviated from 1∶1. A significant increase in prediction error was observed when the mLSTM proportion was excessively high (1∶2, 1∶4). While configurations with a higher sLSTM proportion (2∶1, 4∶1) still outperformed those dominated by mLSTM, their accuracy remained inferior to the balanced xLSTM (1∶1).

The physical characteristics of excitation current in SMs include nonlinear saturation effects under high load and parameter drift due to winding resistance changes with temperature. The sLSTM component is well-suited for modeling rapid dynamic transients. The mLSTM component excels at learning complex nonlinear saturation and coupling effects, utilizing its matrix memory to store inter-variable relationships. The 1∶1 ratio physically corresponds to a balanced modeling of both dynamic response and nonlinearity, which aligns well with the operational process of SMs.

From the perspectives of modeling and training, the balanced ratio achieves a trade-off between structural risk minimization and optimized gradient flow. The gradient stability of sLSTM stems from its normalized gating mechanisms. In contrast, the gradient expressiveness of mLSTM originates from its matrix outer product updates, which empower it to learn complex mappings but also introduce a higher risk of gradient explosion. During backpropagation, the gradients from the two pathways interact and mutually regulate each other.

From an algorithmic principle standpoint, sLSTM and mLSTM exhibit functional complementarity. The former employs a vectorized memory state with element-wise normalized updates, making it suitable for modeling the dynamics of excitation current under steady-state or slowly varying conditions. The latter utilizes a matrix memory and outer product update mechanism, specializing in capturing complex nonlinear effects in current variations.

Predicting excitation current is a hybrid modeling task requiring both temporal dependency learning and feature coupling modeling. The 1∶1 design achieves a balanced fusion of the two modules, preventing either mechanism from dominating and weakening the capabilities of the other.

Furthermore, the results from both comparative and ablation experiments provide supporting evidence that the xLSTM architecture possesses a theoretical foundation for handling both steady-state and dynamic processes. The sLSTM module, through its exponential gating and normalization mechanisms, can stably learn the slowly changing system inertia and also respond rapidly to input disturbances via its gating control. The mLSTM module, with its matrix memory and outer product update rule, can adapt to the intricate coupling relationships among multiple variables during transient states. The dual capability provides xLSTM with a solid basis for learning the mappings among current, voltage, and speed.

4 Conclusions

This study proposes an excitation current prediction model based on the xLSTM network. It addresses the challenge of accurately predicting the excitation current of synchronous machines under multiple operating conditions. The model integrates sLSTM and mLSTM modules to form a balanced hybrid architecture. This design learns long-term dependencies while enabling high-capacity associative memory. Evaluated on operational data from a synchronous machine under different load conditions, the developed xLSTM demonstrates favorable prediction accuracy, generalization capability, and low computational overhead.

Compared to standard LSTM and GRU models, xLSTM shows better performance across error metrics such as RMSE, MAE, and R^2. It also achieves faster computation and higher stability. Ablation studies reveal that a 1∶1 blending ratio of sLSTM to mLSTM yields the best balance between dynamic response and nonlinear coupling modeling. Performance degrades when over-biasing either module. It structurally confirms xLSTM's advantage in modeling physical quantities like excitation current, which exhibit multi-scale dynamics and non-stationary characteristics.

The main contributions of this work are as follows:

(1) An xLSTM architecture suitable for excitation current prediction in synchronous machines is proposed, offering a novel network design approach for intelligent motor modeling.

(2) From a physics-informed perspective, the theoretical rationale for applying xLSTM to excitation current dynamics is explained, aligning network mechanisms with physical behaviors.

(3) Through comparative experiments and ablation analysis, the comprehensive superiority of xLSTM in accuracy, speed, and stability is validated.

Future research will focus on the following three directions:

(1) Constructing or collecting synchronous machine datasets that include dynamic processes such as frequent start-stop operations and variable-speed cutting. An online prediction system based on xLSTM will be developed to further test its capability under transient conditions.

(2) Leveraging multi-physics coupled data to model different key physical quantities, such as excitation current and rotor temperature.

(3) Evaluating the performance of various xLSTM modules and network layer architectures for excitation current prediction across synchronous machines of different power ratings.

References

[1] BURLIKOWSKI W, KIELAN P, KOWALIK Z. Synchronous reluctance machine drive control with fast prototyping card implementation[J]. Archives of Electrical Engineering, 2020, 69(4): 757-769.

[2] CASTAGNARO E, BIANCHI N. High-speed synchronous reluctance motor for electric-spindle application[C]//Proceedings of 2020 International Conference on Electrical Machines (ICEM). Gothenburg, Sweden: IEEE, 2020: 2419-2425.

[3] LIANG J, XU H, LIU W. Design of electric spindle control system for high-grade CNC machine tools[C]//Proceedings of 2023 IEEE International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC). Venice, Italy: IEEE, 2023: 1-5.

[4] XI T, KEHNE S, FEY M, et al. Application of optimal control for synchronous reluctance machines in feed drives of machine tools[C]//Proceedings of 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET). Prague, Czech Republic: IEEE, 2022: 1-6.

[5] SU P, SUN W, YAO X. Analysis of excitation current in synchronous motors based on multiple machine learning models[C]//Proceedings of 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA). Dalian, China: IEEE, 2024: 192-198.

[6] BROSCH A, WALLSCHEID O, BOCKER J. Torque and inductances estimation for finite model predictive control of highly utilized permanent magnet synchronous motors[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8080-8091.

[7] WANG C C, ZHANG J H, JIN L, et al. Rotor displacement control of bearingless switched reluctance motor with single-winding stator[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024, 41(3): 289-299.

[8] FANG Q, YE Z, CHEN C. A review of power transformer vibration and noise caused by silicon steel magnetostriction[J]. Electronics, 2024, 13(5): 968.

[9] HAN Z, HUANG R, HUANG B, et al. Data-driven and knowledge-guided approach for NC machining process planning[J]. Computer-Aided Design, 2023, 162: 103562.

[10] SINDLER Y, LINEYKIN S. Static, dynamic, and signal-to-noise analysis of a solid-state magnetoelectric (Me) sensor with a spice-based circuit simulator[J]. Sensors, 2022, 22(15): 5514.

[11] MENG X Y, CAO X, SHI R J, et al. Fast axial positioning with linear-velocity control for linear-rotary switched reluctance motor[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024, 41(3): 300-310.

[12] WANG X, LU H, HUANG X, et al. Three-dimensional time-varying sliding mode guidance law against maneuvering targets with terminal angle constraint[J]. Chinese Journal of Aeronautics, 2022, 35(7): 303-319.

[13] KIM S, PETRUNIN I, SHIN H-S. A review of Kalman filter with artificial intelligence techniques[C]//Proceedings of 2022 Integrated Communication, Navigation and Surveillance Conference (ICNS). Dulles, VA, USA: IEEE, 2022: 1-12.

[14] SUN X, WEN C, WEN T. Maximum correntropy

high-order extended Kalman filter[J]. Chinese Journal of Electronics, 2022, 31(1): 190-198.
[15] ZHAO D, LI B, LU F, et al. Deep bilinear koopman model predictive control for nonlinear dynamical systems[J]. IEEE Transactions on Industrial Electronics, 2024, 71(2): 16077-16086.
[16] USAMA M, KIM J. Model-free current control solution employing intelligent control for enhanced motor drive performance[J]. Scientific Reports, 2025, 15: 60.
[17] JI Z Y, ZHANG J J, LIU Z H, et al. A CNN-based method for sparse SAR target classification with grad-CAM interpretation[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2025, 42(4): 525-540.
[18] ZENG X, MA Y, WAN J, et al. A fusion modeling approach for six-phase hybrid excitation synchronous motor, leveraging finite-element analysis, and experimental data-driven[J]. IEEE Access, 2025, 13: 32781-32793.
[19] LU Y, HUANG K, WANG B, et al. Data-driven modeling and compensation strategy of PMSM considering core loss and saturation[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2024, 12(2): 1894-1905.
[20] CICEO S, CHAUVICOURT F, GYSELINCK J, et al. Data-driven electrical machines structural model using the vibration synthesis method[J]. IEEE Transactions on Transportation Electrification, 2022, 8(3): 3771-3781.
[21] SUN Z, WANG Q, QIAN Z, et al. Estimate end-winding temperature for water-cooled interior permanent magnet motors under variable-speed driving cycles: A solution to the exact challenge[J]. IEEE Transactions on Transportation Electrification, 2023, 9(1): 370-381.
[22] ZHANG Z, HU Y, LUO G, et al. An embedded fault-tolerant control method for single open-switch faults in standard PMSM drives[J]. IEEE Transactions on Power Electronics, 2022, 37(7): 8476-8487.
[23] GUILLÉN C E, DE PORRAS COSANO A M, TIAN P, et al. Synchronous machines field winding turn-to-turn fault severity estimation through machine learning regression algorithms[J]. IEEE Transactions on Energy Conversion, 2022, 37(3): 2227-2235.
[24] GLŬČINA M, ANĐELIĆ N, LORENCIN I, et al. Estimation of excitation current of a synchronous machine using machine learning methods[J]. Computers, 2022, 12(1): 1.
[25] DUAN J J, LU Z, DU Z Q. Attention-based multi-scale CNN and LSTM model for remaining useful life estimation[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2025, 42(S): 64-77.
[26] BECK M, PÖPPEL K, SPANRING M, et al. xLSTM: Extended long short-term memory[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada: Curran Associates Inc, 2024: 107547-107603.
[27] KAHRAMAN H T, BAYINDIR R, SAGIROGLU S. A new approach to predict the excitation current and parameter weightings of synchronous machines based on genetic algorithm-based K-NN estimator[J]. Energy Conversion and Management, 2012, 64: 129-138.
[28] KAHRAMAN H T. Metaheuristic linear modeling technique for estimating the excitation current of a synchronous motor[J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2014, 22: 1637-1652.
[29] ZARZYCKI K, ŁAWRYŃCZUK M. Advanced predictive control for GRU and LSTM networks[J]. Information Sciences, 2022, 616: 229-254.
[30] LI W, LAW K L E. Deep learning models for time series forecasting: A review[J]. IEEE Access, 2024, 12: 92306-92327.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Nos. 52575557, 52375140) and the Jiangxi Key Laboratory of High-End CNC Machine Tools.

Authors

The first author Mr. **CHE Zhongyuan** received the M.S. degree in intelligent manufacturing engineering from School of Mechanical Engineering and Automation, Beihang University, in 2024. Currently, he is pursuing the Ph.D. degree at Beihang University. His research interests include prediction of machining processes, artificial intelligence methods, and advanced manufacturing modeling.

The corresponding author Prof. **PENG Chong** is a tenured professor and Ph.D. supervisor at School of Mechanical Engineering and Automation, Beihang University. His research primarily focuses on the reliability of machine tools, modeling of smart manufacturing, optimization of manufacturing processes, and manufacturing service systems. His work aims to develop next-generation and efficient manufacturing paradigms for high-end CNC machine tools and high-

performance CNC systems.

Author contributions Mr. CHE Zhongyuan proposed the methodology, interpreted the conceptual framework, developed the software code, and drafted and revised the manuscript. Prof. PENG Chong supervised the research execution, critically reviewed the manuscript content, and provided constructive suggestions to refine the study design. Ms. ZHANG Rui contributed to language editing, manuscript polishing, and assisted in research implementation. Mr. WANG Chi provided feedback on figures and tables and assisted in manuscript revisions. All authors reviewed and approved the final version of the manuscript.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

面向电主轴驱动的同步电机励磁电流xLSTM预测方法

车众元[1,2,3], 彭 翀[1,2,3], 张 蕤[1], 王 弛[1]

(1.北京航空航天大学机械工程及自动化学院,北京 102206,中国; 2.北京航空航天大学江西研究院,南昌 330096,中国; 3.北京航空航天大学江西研究院高端数控机床江西省重点实验室,南昌 330096,中国)

摘要: 同步电机广泛应用于电主轴等工业驱动场景,励磁电流的准确建模和预测是决定同步电机(Synchronous machines, SMs)高性能控制的关键因素之一。然而,励磁电流受温漂、磁饱和等耦合非线性因素影响,难以在多工况范围内实现兼具高精度与泛化的预测。为此,本文提出一种基于扩展长短期记忆(Extended long short-term memory, xLSTM)网络的励磁电流预测模型,该模型融合标量长短期记忆单元(scalar LSTM, sLSTM)与矩阵长短期记忆单元(matrix LSTM, mLSTM),利用指数门控机制增强对复杂非线性映射及长期动态依赖关系的学习能力。其中,sLSTM的向量并行记忆结构适合理解温漂引起的缓慢参数变化,mLSTM的矩阵联想记忆机制擅长学习磁饱和等多变量非线性耦合效应,二者形成互补混合架构。通过覆盖可变负载与多励磁状态的同步电机监测数据,开展实验并对比测试所提模型与经典长短期记忆网络、门控循环单元等基准模型的预测效果。此外,设计不同比例的标量与矩阵长短期记忆单元进行了消融实验。多项误差指标与计算时间的综合评估结果表明,所提出的扩展长短期记忆网络模型在预测精度、泛化能力、计算效率及稳定性方面均更具优势。最后,本文从网络架构与算法机理层面讨论和分析了其性能提升的原因。本研究为同步电机励磁电流的建模提供了一种新的数据驱动思路,对高保真电机状态估计等应用场景具有参考价值。

关键词: 同步电机;励磁电流建模;数据驱动方法;扩展长短期记忆网络

研究亮点:

1. 提出基于扩展长短期记忆网络(xLSTM)励磁电流预测模型实现非线性映射与依赖学习。
2. 从物理机理层面阐释xLSTM双模块结构(sLSTM和mLSTM)与励磁电流动态特性对应关系。
3. 对比和消融实验测试了sLSTM和mLSTM不同比例对精度、泛化、效率及稳定性的影响。
4. 为工业驱动场景高保真电机状态估计提供一种低开销与高泛化的数据驱动建模思路。