Friction
https://mc03.manuscriptcentral.com/friction
https://doi.org/10.26599/FRICT.2026.9441313

Research Article

A cU-Net-based surrogate model for predicting rough surface contact parameters in line-contact friction

Yichun Xia, Hui Cao, Yonggang Meng ✉

State Key Laboratory of Tribology in Advanced Equipment (SKLT), Tsinghua University, Beijing, 100084, China

✉ Corresponding author. E-mail: mengyg@mail.tsinghua.edu.cn

Received: May 12, 2026; Revised: July 17, 2026; Accepted: September 6, 2026

Abstract: Accurate and efficient prediction of contact parameters on rough surfaces is critical for the tribological design, condition monitoring, and failure analysis of line-contact components. Although physics-based models offer high prediction fidelity, their multi-field coupled iterative solvers are computationally expensive and complex, limiting practical engineering deployment. This work proposes a digital surrogate model based on the conditional U-Net (cU-Net) for predicting the spatial distributions of contact parameters, including contact temperature, pressure, and heat flux, on rough surfaces in line-contact friction. Real surface topographies are augmented via the Generative Patch Nearest-Neighbor (GPNN) model, and a wide-ranging training dataset is constructed through batch computation using a boundary lubrication physics-based model. The cU-Net incorporates an operating condition embedding mechanism to enable unified prediction across multiple operating conditions. Multi-scale spatial and channel attention modules are further integrated to enhance cross-scale feature extraction. A four-component progressive composite loss, assembled from established loss terms and tailored to the sparse extreme-value characteristics of rough surface contact parameters, is combined with Bayesian hyperparameter optimization and a three-stage progressive training strategy. On completely unseen rough surfaces the model achieves mean relative errors of 3.53%, 12.32%, and 13.66% for the temperature, pressure, and heat flux fields, respectively, with peak and top-5% errors that are lower still, and it is approximately 10^5 times faster than the physics-based model on the same workstation. Out-of-distribution tests quantify the degradation of accuracy under extrapolated operating conditions. Retraining with datasets generated from the corresponding physics-based models is expected to extend the framework to other lubrication regimes.

Keywords: line-contact friction; contact parameters; digital surrogate model; rough surface

1 Introduction

Line-contact friction is prevalent in critical industrial components such as gears[1, 2], cylindrical roller bearings[3], and cams. During the friction process, contact pressure, contact temperature, and lubricant film thickness are key parameters characterizing the physical state of the contact area. Contact pressure governs the stress distribution within the contact zone and directly influences the initiation and evolution of surface damage such as fatigue and wear[4, 5]. Contact temperature affects lubricant viscosity and material properties, and excessive temperature can lead to lubrication failure and surface damage, such as seizure [6, 7]. Lubricant film thickness directly determines the lubrication regime and friction performance[8, 9]. These parameters are highly coupled, and their spatial distributions serve as critical references for tribological design, condition monitoring, and service life prediction. Therefore, accurate and efficient acquisition of the spatial distributions of contact parameters during friction is of great significance for the optimal design, reliable operation, and intelligent maintenance of line-contact tribological pairs.

Contact parameters during friction are difficult to measure directly through experimental means. Although partial information can be obtained using techniques such as thermocouples and infrared thermography for temperature measurement[10, 11], and optical interferometry and capacitance methods for lubricant film thickness[12, 13], existing measurement techniques remain fundamentally limited in spatiotemporal resolution[14, 15]. The extremely small size, optical inaccessibility, and high-speed motion of the contact zone make it practically impossible to capture the transient variations and local distributions of contact parameters. As a result, measured data predominantly reflect bulk substrate parameters rather than the true physical state within the contact area. Therefore, establishing physics-based models for the quantitative prediction of contact parameters has become an important alternative approach.

Physics-based models predict the spatial distributions of contact parameters such as pressure, temperature, and film thickness by solving governing equations of contact mechanics, lubrication, and heat transfer. Well-established computational frameworks, including elastohydrodynamic lubrication (EHL) models[16, 17], rough-surface contact models[18, 19], and flash temperature models[20–22], have been widely applied to the analysis of contact parameters in line-contact components. However, these models rely on multi-field coupled iterative solvers that become computationally prohibitive when real rough surfaces are considered, and may suffer from poor convergence and numerical instability under extreme operating conditions. These limitations, combined with the domain expertise required for model operation, present significant barriers to their rapid deployment in engineering practice. Consequently, it is of great importance to develop surrogate models that balance prediction accuracy and computational efficiency, so as to bring contact parameter prediction into practical engineering applications.

Data-driven surrogate models have emerged as a promising alternative for the rapid prediction of contact parameters in friction, leveraging their efficient inference capability. For film thickness prediction, Marian et al. [23] and others[24, 25] established predictive models for central and minimum film thickness under line-contact and point-contact conditions using support vector machines, Gaussian process regression, and multilayer perceptron (MLP). For coefficient of friction (COF) prediction, Wang et al.[26] developed a surrogate model based on thermal elastohydrodynamic lubrication (TEHL) model, achieving high prediction accuracy for both COF and film thickness in face-gear drives. Wang et al.[27] employed CFD simulations to generate training data and compared multiple machine learning methods for friction prediction in textured journal bearings, with ANN achieving the highest prediction accuracy among the methods evaluated. For temperature prediction, Singh et al.[28] compared several machine learning algorithms for predicting the maximum flash temperature in TEHL model. Prajapati et al.[29] predicted the asperity load ratio for rough surface contacts using surface topography statistics and operating conditions as inputs, while Kalliorinne et al.[30] and Suman and Prajapati [31] employed ANN to predict real contact area and related contact mechanical responses for rough surfaces. These studies achieved high prediction accuracy within their respective operating condition ranges. Multiple comparative studies have demonstrated the comprehensive advantages of MLPs in prediction accuracy and inference speed[23, 25, 28]. The common limitation, however, is that all of them treat single scalar quantities as prediction targets, and are therefore unable to provide spatial distribution information of physical fields within the contact zone.

Some studies have extended the prediction target from scalar quantities to spatial field distributions. For pressure and film thickness field prediction, Hess and Shang[32] employed a U-Net to predict the full-field EHL pressure distribution in sliding bearings, and Kelley et al.[33] extended the prediction target from scalar film thickness to one-dimensional pressure and film thickness distributions. For temperature field prediction, Cartwright et al.[34] achieved full-field prediction of pressure, film thickness, and temperature distributions in journal bearings across a range of operating conditions, while Wu et al.[35] reconstructed the full-field temperature distribution of wet clutches from sparse sensor measurements using a self-attention U-Net. For rough surface contact, Jang and Jang[36] employed a U-Net with surface height maps as direct inputs to predict the real contact area distribution on rough surfaces, achieving a computational efficiency improvement of over 95% compared to conventional numerical methods. However, the above studies share two common limitations. First, most full-field prediction studies address smooth surface contact problems, in which the spatial distributions of physical quantities are relatively regular and continuous, and do not exhibit the localized extreme values and sparse distribution characteristics arising from asperity contacts on real rough surfaces; even the rough surface study of Jang and Jang[36] is limited to contact area distribution and does not extend to physical quantities such as contact pressure and flash temperature. Second, these models are typically trained and validated under fixed or narrowly defined operating conditions, lacking an explicit mechanism to incorporate operating condition parameters, and are therefore unable to generalize across varying tribological conditions.

To address the above limitations, this study proposes a cU-Net (conditional U-Net) -based digital surrogate model for full-field prediction of contact parameters on rough surfaces in line-contact friction. The main contributions of this work are as follows:

(1) Real rough surface topographies were experimentally measured and augmented using the Generative Patch Nearest-Neighbor (GPNN) model to generate a large and diverse set of rough surfaces. Combined with an established boundary lubrication physics-based model, contact parameter distributions were obtained across a wide range of operating conditions, forming a large-scale training dataset that provides a reliable data foundation for surrogate model development.

(2) A cU-Net architecture is proposed, in which operating condition parameters are incorporated into the model to enable unified prediction across multiple operating conditions. Multi-scale spatial

attention modules are further introduced to enhance the model's capability in capturing the spatial features of physical field distributions.

(3) A four-component composite loss function is designed to simultaneously improve prediction accuracy, preserve spatial continuity, and maintain frequency-domain fidelity for physical fields with sparse extreme-value characteristics on rough surfaces. A progressive training strategy is further adopted to ensure stable convergence.

2 Model methodology

2.1 Model framework

The physical problem addressed by the underlying physics-based model is stated first. It solves a three-dimensional rough-surface contact problem on the nominal Hertzian contact domain, coupling Boussinesq elastic contact, frictional heat generation at the asperities and three-dimensional heat conduction, and it outputs the in-plane full-field distributions of contact pressure, temperature and heat flux. The formulation and its numerical implementation are given in Ref.[37]. The model is restricted to the boundary lubrication regime, in which the load is predominantly carried by solid–solid asperity contacts, so that hydrodynamic and mixed lubrication effects are neglected and the heat is assumed to originate mainly from friction at the asperity contacts within the real contact area. Lubricant convective cooling, interfacial thermal contact resistance, thermal expansion, and phase-change or tribochemical phenomena are not explicitly accounted for. The surrogate model developed below is a fast approximation of this specific solver, and its validity is bounded by these assumptions.

The surrogate model consists of six sequential stages: data generation, data preprocessing, data splitting, model construction, hyperparameter optimization, and model training and validation.

In the data generation stage, real surface topographies are acquired using an optical 3D surface profiler as reference samples. A surface generation model is then applied to augment these measurements, producing a large collection of synthetic rough surfaces that share similar statistical characteristics but differ in local morphological details. The generated surfaces are paired with varying operating conditions (including load, entrainment velocity, slide-to-roll ratio, and others) and fed into a boundary lubrication physics-based model. The spatial distributions of contact parameters such as contact pressure and temperature are then systematically computed through batch calculation, constructing a comprehensive training dataset.

In the data preprocessing stage, appropriate normalization is applied to each physical quantity to address the significant statistical non-uniformity in their distributions, improving the model's balanced learning capability. In the data splitting stage, the complete dataset is divided into training, validation, and test sets in a 7:2:1 ratio, used respectively for model parameter learning, hyperparameter tuning and overfitting monitoring, and independent evaluation of generalization performance.

In the model construction stage, a cU-Net is built upon the U-Net architecture with the introduction of a condition embedding mechanism, spatial and channel attention modules for enhanced multi-scale feature extraction, and a progressive composite loss function tailored to the distributional characteristics of rough-surface contact parameters. In the hyperparameter optimization stage, Bayesian optimization is employed to systematically search for the optimal configuration of key hyperparameters in the cU-Net.

2.2 Rough surface generation model (GPNN)

Real rough surfaces can be obtained through experimental measurement, but the number of measurable samples is limited and insufficient to meet the large-scale data requirements of deep learning models. Established random-rough-surface generators, such as fractal models[38] and spectrum-targeting random-process methods[39], are designed to reproduce a prescribed power spectral density (PSD) across scales, which is the physically meaningful fingerprint of a rough surface in the standard contact-mechanics framework. They are not adopted here because the measured surfaces are markedly non-Gaussian and plateau-bearing, and a purely spectral generator does not constrain the height distribution or the local plateau–valley morphology. The Generative Patch Nearest-Neighbor (GPNN) model[40] is adopted instead, because it reproduces both the spectral content and the measured height statistics, as verified in Section 3.1. This is a choice of which surface descriptors to preserve, and not a claim of superiority over spectrum-targeting generators. In this study, experimentally measured surface topographies serve as the initial reference samples, which are then expanded into a diverse rough-surface dataset using the GPNN model.

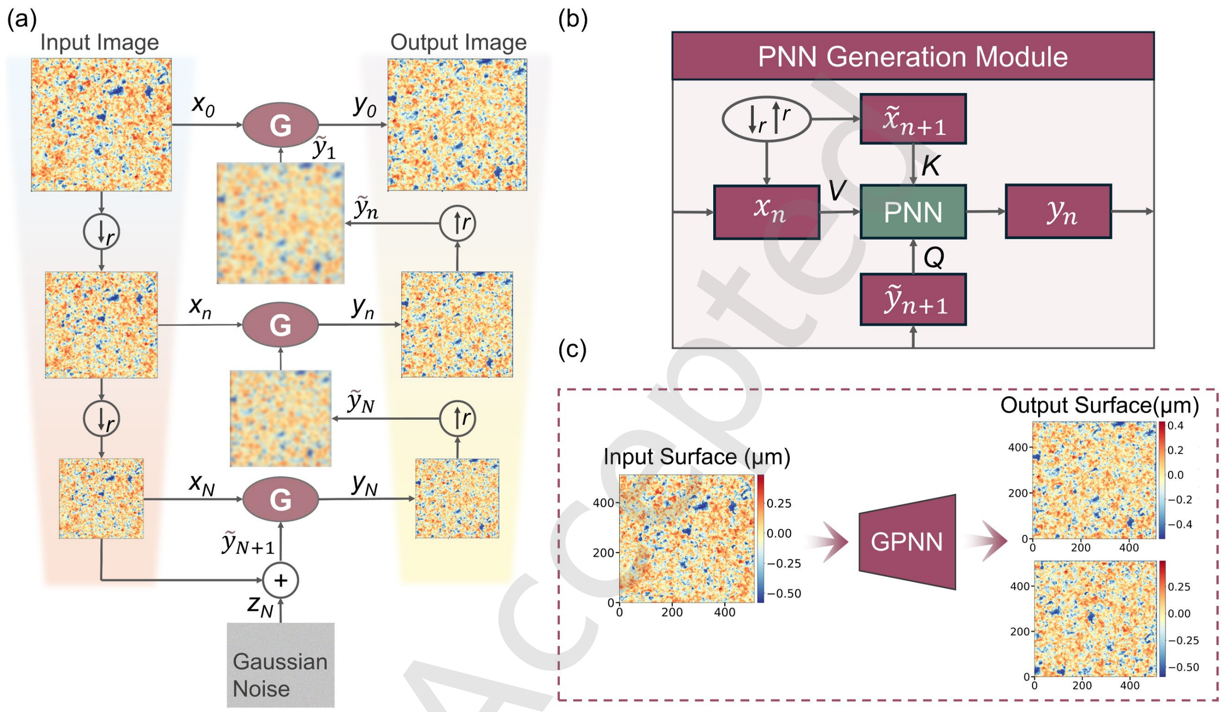

Fig. 1 The GPNN model (a) Overall architecture (b) PNN generation module (c) Illustration of generated rough surfaces.

As shown in Fig. 1(a), GPNN employs a coarse-to-fine multi-scale generation strategy. At coarse scales, the global structure and large-scale height distribution are established, while finer scales progressively incorporate local asperity details and short-wavelength roughness components, enabling a gradual refinement from macroscopic morphology to microscale surface texture.

At each scale, GPNN operates directly on height maps through two decoupled steps, matching and replacement, as illustrated in Fig. 1(b). In the matching step, the current height estimate is low-pass filtered and divided into overlapping local patches. Each such patch is compared with the patches of an identically filtered version of the measured reference height map, and the reference patch that minimizes a normalized sum of squared height differences is selected; this is what is meant by matching. In the replacement step, the corresponding patch of the original, unfiltered reference surface is pasted into the synthesized surface, so that the fine-scale roughness of the measured topography is transferred to the generated one. The normalization of the similarity score prevents a small number of reference patches from being reused everywhere, so that all regions of the reference surface contribute to the output and the global height distribution and spectral characteristics of the measurement are preserved. Since the patches overlap spatially, the final height map is composited by Gaussian-weighted averaging, which suppresses discontinuities at the patch boundaries. The matching-and-replacement process is repeated several times at each scale to refine the generated surface progressively. The rough surfaces generated by GPNN are consistent with the original measured samples in the key height statistics and in the power spectral density, as quantified in Section 3.1, while local morphological variations are introduced through the inherent randomness of the patch matching process, providing sufficiently diverse input data for cU-Net model training.

2.3 cU-Net model

Predicting the full-field distributions of rough-surface contact parameters poses several challenges. These include handling two-dimensional spatial inputs, extracting cross-scale features across global contact morphology and localized asperity extremes, and embedding operating conditions for multi-condition prediction.

Considering these requirements, cU-Net, built upon the U-Net architecture[41], is selected as the backbone architecture of the surrogate model. Compared to MLP, which struggles with two-dimensional spatial structures, and Transformer, whose global self-attention lacks the locality prior required by asperity-scale contact and is data-hungry at the present dataset scale, cU-Net offers a better balance between modeling capacity and computational efficiency for the present task.

cU-Net is a condition-embedded encoder-decoder deep learning model comprising four components: encoding blocks, a bottleneck layer, decoding blocks, and skip connections, as illustrated in Fig. 2.

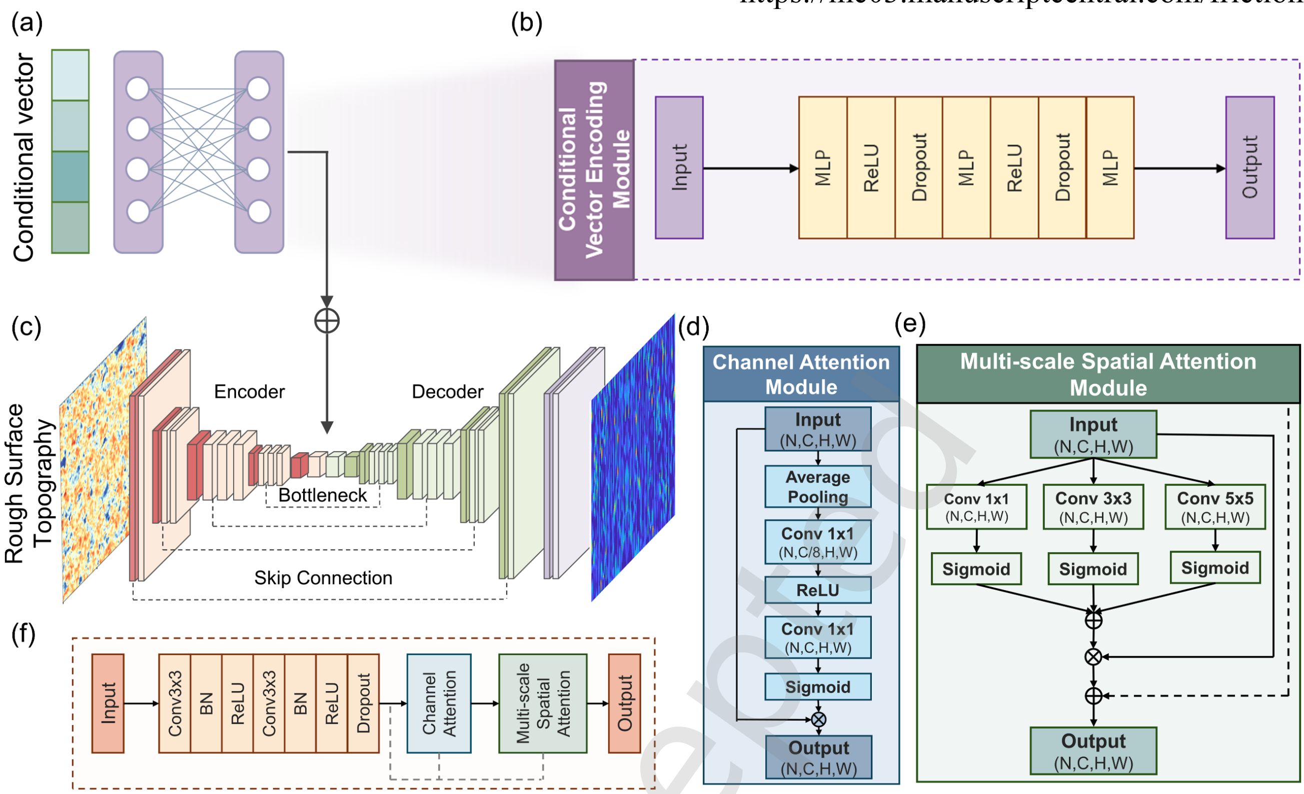

Fig. 2 cU-Net model (a) Condition vector encoding (b) Condition vector encoding module (c) Overall architecture of cU-Net (d) Channel attention module (e) Multi-scale spatial attention module (f) Structure of the U-Net encoding and decoding blocks.

The rough surface topography is represented as a single-channel discretized two-dimensional scalar height map and processed by a convolutional encoder to extract cross-scale spatial features layer by layer. Both the input height map and the corresponding target physical fields are defined on the same non-dimensional computational domain spanning the nominal Hertzian contact region. Therefore, the network learns the mapping between non-dimensional fields, while the corresponding physical dimensions are recovered through the inverse scaling after prediction. As shown in Fig. 2f, each encoder module comprises two 3×3 convolutional layers each followed by batch normalization and ReLU activation, a Dropout regularization layer, and sequential channel attention and multi-scale spatial attention modules.

Operating condition parameters, including load, velocity, and geometric parameters, are encoded as a one-dimensional condition vector and mapped to a high-dimensional representation through a three-layer perceptron, as illustrated in Fig. 2a and b. The encoded condition information is then fused with the rough surface features at the network bottleneck via a broadcasting mechanism:

$$\mathbf{F}_{bottleneck} = \mathbf{F}_{d}^{enc} + \mathbf{c}_{emb} \otimes \mathbf{1}_{H_d \times W_d} \tag{1}$$

where $\mathbf{F}_{d}^{enc} \in \mathbb{R}^{C_d \times H_d \times W_d}$ is the feature map at the deepest encoder layer, C_d denotes the number of channels, $H_d \times W_d$ is the corresponding spatial dimension, $\mathbf{c}_{emb}$ is the condition embedding vector obtained by encoding the operating condition parameters through the perceptron, $\otimes$ denotes the tensor outer product operation, $\mathbf{1}_{H_d \times W_d}$ is an all-ones matrix. This fusion mechanism enables the network to modulate its response to rough surface features according to different operating conditions, allowing unified prediction across multiple conditions within a single model. The decoder restores spatial resolution through bilinear up-sampling, while skip connections directly transfer shallow encoder features to the corresponding decoder layers, preserving multi-scale spatial details of the rough surface and facilitating the recovery of fine-grained structures in the output physical fields.

To enhance the network's sensitivity to key features of the contact parameter fields, an attention module integrating channel attention (Fig. 2d) and multi-scale spatial attention (Fig. 2e) is incorporated. The channel attention mechanism captures inter-channel dependencies through global average pooling, adaptively recalibrating the importance of different feature channels[42]:

$$\mathbf{A}_{c} = \sigma\left(W_{c}^{2} \cdot \mathrm{ReLU}\left(W_{c}^{1} \cdot \mathrm{GAP}\left(\mathbf{F}\right)\right)\right) \tag{2}$$

where $\mathrm{GAP}\left(\mathbf{F}\right) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\mathbf{F}_{:,h,w} \in \mathbb{R}^{C}$ denotes global average pooling, H and W are the spatial height and width of the feature map, C is the number of channels, $W_{c}^{1} \in \mathbb{R}^{C/r \times C}$ and $W_{c}^{2} \in \mathbb{R}^{C \times C/r}$ are trainable weight matrices of the dimensionality reduction and expansion weight matrices, respectively, with r denoting the compression ratio, $\sigma\left(\cdot\right)$ is the Sigmoid activation function.

The multi-scale spatial attention module employs three parallel convolutional branches with kernel sizes of 1×1, 3×3, and 5×5 to extract spatial features at different scales, which are then aggregated through soft max-normalized learnable weights[43, 44]:

$$\mathbf{A}_s = \sum_{k\in\{1,3,5\}} \beta_k \cdot \sigma\left(\mathrm{Conv}_{k\times k}\left(\mathbf{F}\right)\right) \tag{3}$$

where $\mathrm{Conv}_{k\times k}\left(\cdot\right)$ denotes a convolution operation with kernel size *k*×*k*, β_k is the soft max-normalized fusion weight. This design enables the network to simultaneously capture localized asperity details and global contact zone structural information, with a residual connection applied to the final output to stabilize the training process.

To handle the sparse extreme-value characteristics of rough-surface contact parameters, a four-component composite loss function is designed:

$$L_{total} = \lambda_1 L_{MAE} + \lambda_2 L_{grad} + \lambda_3 L_{focal} + \lambda_4 L_{spectral} \tag{4}$$

where λ_i is the weighting coefficient for each loss component, L_{MAE} is the mean absolute error loss, L_{grad} is the adaptive gradient loss, L_{focal} is the Focal loss, $L_{spectral}$ is the spectral loss.

The MAE loss is a standard L1 reconstruction term, adopted here unchanged to provide robust baseline supervision:

$$L_{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{5}$$

where y_i and $\hat{y}_i$ are the ground truth and predicted values of the physical quantity at pixel *i*, respectively, $N = H \times W$ is the total number of pixels.

The adaptive gradient loss constrains the spatial gradients of contact parameters, imposing stronger penalties in high-gradient regions to ensure the continuity and physical plausibility of the predicted contact parameter distributions:

$$L_{grad} = \frac{1}{2N}\sum_{i=1}^{N}\left[\alpha_{x,i}\cdot\left(G_x\left(y_i\right) - G_x\left(\hat{y}_i\right)\right)^2 + \alpha_{y,i}\cdot\left(G_y\left(y_i\right) - G_y\left(\hat{y}_i\right)\right)^2\right] \tag{6}$$

where $G_x(\cdot)$ and $G_y(\cdot)$ are Sobel operators that extract spatial gradients of the entire physical field in the *x* and *y* directions, respectively, with the subscript *i* denoting the value at pixel *i*, $\alpha_{x,i}$ and $\alpha_{y,i}$ are adaptive weights dynamically adjusted according to the gradient magnitudes in the *x* and *y* directions at each location.

The Focal loss[45] reinforces learning on hard samples with large prediction errors, directing the network's attention toward extreme-value regions of rough-surface contact parameters:

$$L_{focal} = \frac{1}{N}\sum_{i=1}^{N}\left(1-\exp\left(-\left(y_i-\hat{y}_i\right)^2\right)\right)^{\gamma}\cdot\left(y_i-\hat{y}_i\right)^2\cdot m_i \tag{7}$$

where γ is the focusing parameter that increases the relative weight of hard samples with large prediction errors, m_i is the sample weight mask that assigns higher weights to high-value regions.

Pixel-wise losses are known to under-resolve the high-wavenumber content of reconstructed physical fields, as documented in the super-resolution of turbulent flows[46, 47]. The established remedy is a frequency-domain penalty[48, 49], which is adopted here in its standard form as the L1 discrepancy between the amplitude spectra of the predicted and reference fields. It prevents the network from either over-smoothing the asperity-scale texture or introducing non-physical high-frequency oscillations:

$$L_{spectral} = \frac{1}{N}\sum_{u=1}^{H}\sum_{v=1}^{W}\left|\mathcal{F}(y)_{u,v}-\mathcal{F}(\hat{y})_{u,v}\right|^2 \tag{8}$$

where $\mathcal{F}(\cdot)$ denotes the amplitude spectrum of the two-dimensional fast Fourier transform, (u,v) are the frequency domain coordinates.

3 Dataset Construction and Model Training

3.1 Data generation and preprocessing

Fig. 3 presents an overview of the data generation process for contact parameters under boundary lubrication conditions. The input data consist of two components: operating condition parameters and rough surface topography data. The operating condition parameters include key variables such as load, entrainment velocity, and slide-to-roll ratio, with their ranges summarized in Table 1. A constant COF of 0.15 is adopted throughout all simulations.

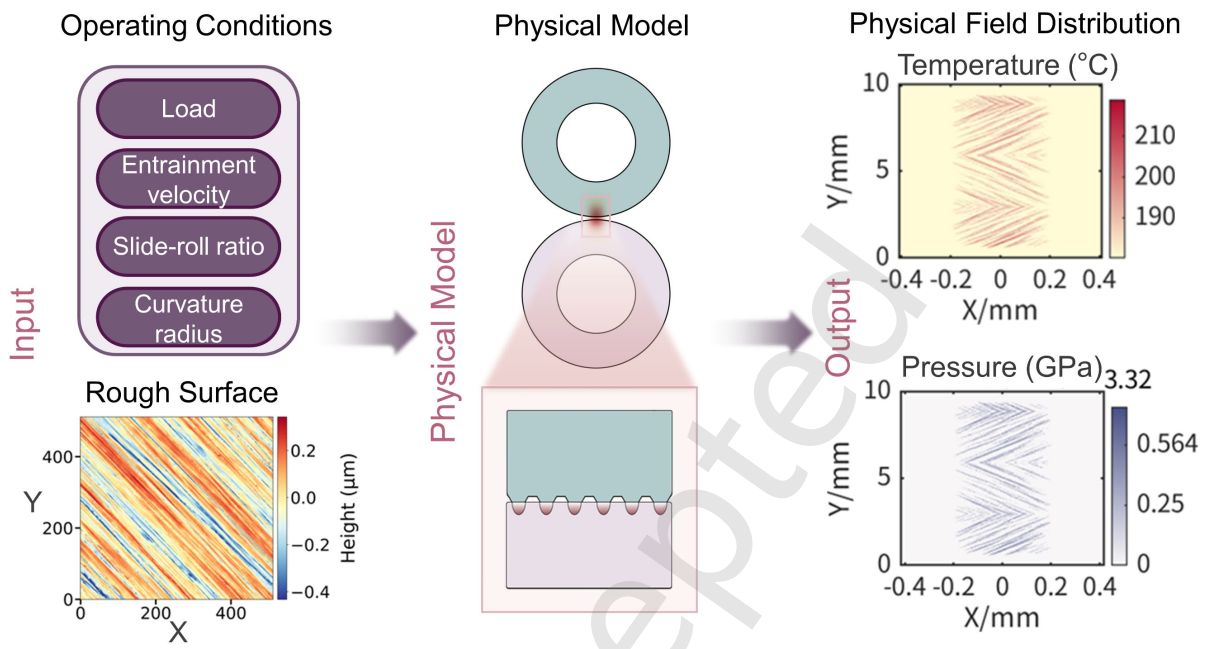

Fig. 3 Data generation process for contact parameters under boundary lubrication conditions.

Table 1 Operating conditions

Parameter	Range	Parameter	Range
Entrainment velocity (m/s)	1~3	Slide-to-roll ratio	0.05~0.5
Load per unit length (N/mm)	10~200	Equivalent radius (mm)	1~9
COF	0.15		

The rough surface data combine experimentally measured topographies with GPNN-based augmentation. To eliminate potential information leakage, a surface-isolated partitioning strategy is adopted before data augmentation. Specifically, the approximately 100 measured rough surfaces are first divided into training, validation, and test subsets at a ratio of 7:2:1. GPNN augmentation is then performed independently within each subset, ensuring that all augmented samples derived from the same measured surface remain in the same subset and that no parent-surface information is shared between the training, validation, and test sets. The augmentation preserves the local asperity morphology of the parent surface while varying the spatial arrangement of that morphology, thereby expanding each reference surface into a family of distinct but statistically consistent realizations. Each augmented surface is compared pairwise with its measured parent using the height moments (Sq, Ssk, Sku), the autocorrelation length (Sal), and the radially averaged absolute power spectral density with its log–log slope β (Table 2, Fig. 4). The measured surfaces are markedly non-Gaussian (Ssk = −0.305 ± 0.579, Sku = 3.695 ± 1.486), and the augmentation reproduces this height distribution (Ssk: R^2 = 0.927, Sku: R^2 = 0.900) as well as the RMS height and the correlation length. As Fig. 4 shows, the augmented and measured spectra agree closely across most of the resolved range. The augmented surfaces are slightly less energetic. They carry about 10 % less spectral power than their parents, the median paired ratio being 0.94 at wavelengths of 21–127 µm and 0.89 at 3.4–20 µm.

Table 2 Statistical parameters of the measured and GPNN-augmented surfaces

Parameter	Measured	GPNN	R^2
Sq	0.146 ± 0.100	0.145 ± 0.099	0.925
Ssk	−0.305 ± 0.579	−0.411 ± 0.599	0.927
Sku	3.695 ± 1.486	4.001 ± 2.305	0.900
Sal	17.3 ± 12.1	16.7 ± 11.9	0.897
PSD slope β	−1.427 ± 0.566	−1.515 ± 0.533	0.932

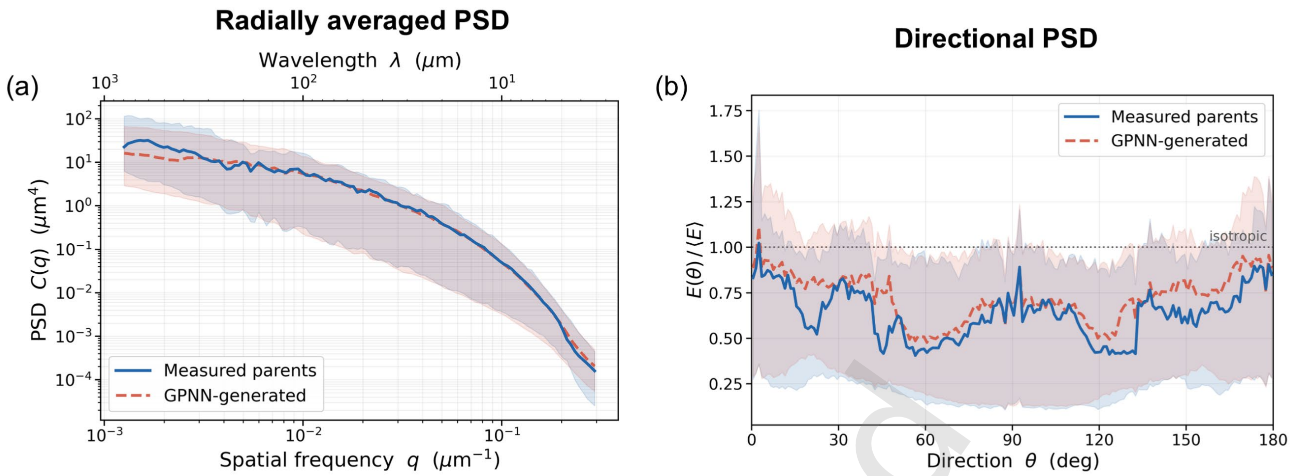

Fig. 4 Spectral validation of the augmented surfaces (a) Radially averaged absolute power spectral density (b) Directional PSD energy.

Various combinations of the augmented surfaces and operating conditions are fed into the physics-based model established in Ref. [37] through which the spatial distributions of contact temperature, contact pressure, and heat flux are systematically computed via batch calculation across a wide range of operating conditions and surface topographies. The detailed theoretical formulation and numerical implementation of the physics-based model can be found in Ref. [37]. A comprehensive training dataset of 35,000 samples is thereby constructed, providing a reliable data foundation for surrogate model development.

The statistical distributions of contact parameters in the dataset are shown in Fig. 5, all of which are consistent with the physical characteristics of rough surface contact. Surface height approximately follows a zero-mean Gaussian distribution. Contact pressure exhibits a pronounced right-skewed distribution, as a small number of highly prominent asperities bear extremely high loads, forming an extended tail. The distribution of flash temperature rise is physically correlated with that of contact pressure. Heat flux rate shows the most pronounced right skewness, reflecting its heightened sensitivity to local contact conditions.

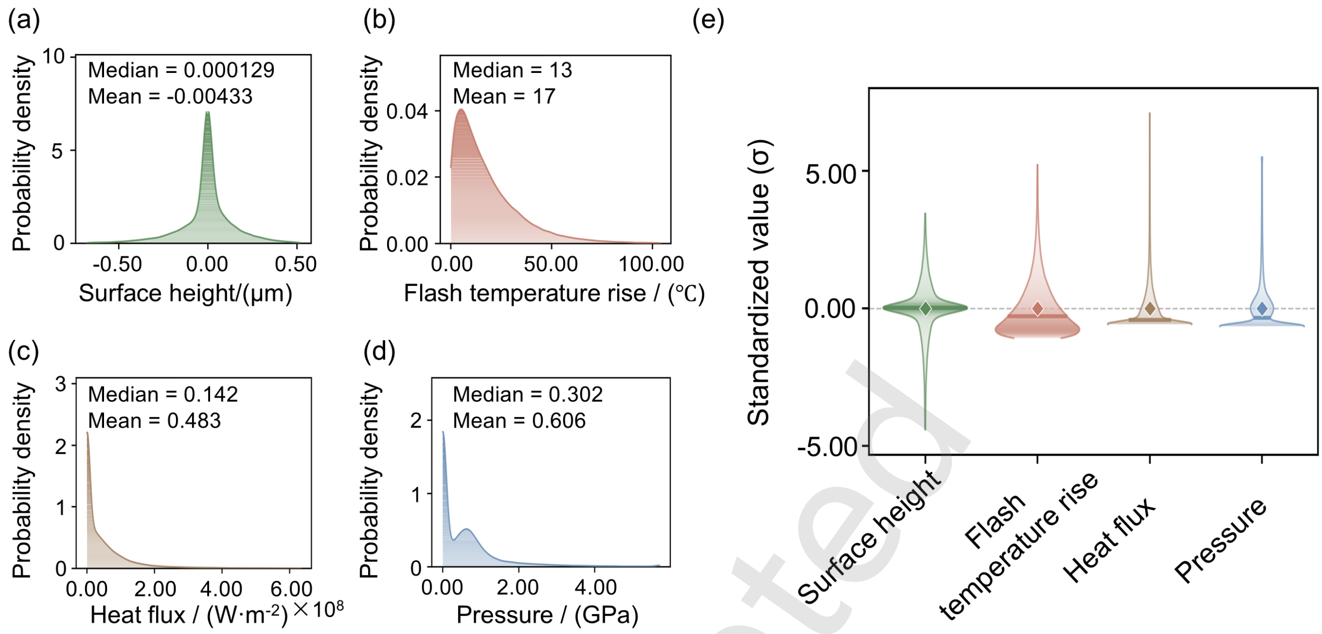

Fig. 5 Statistical distributions of the physical quantities in the dataset (a) Surface height distribution (b) Flash temperature rise distribution (c) Heat flux distribution (d) Contact pressure distribution (e) Data density and dispersion of the rough surface height and three physical fields.

Prior to training, the dataset undergoes preprocessing to meet the requirements of model training. Operating condition parameters, which are uniformly sampled within their prescribed ranges, are processed using standard normalization. Since surface height data contain negative values and are therefore incompatible with logarithmic transformations that require strictly positive inputs, a Yeo-Johnson transformation is applied in conjunction with robust scaling based on the median and interquartile range, correcting the residual deviation between the actual distribution and the ideal Gaussian form. For flash temperature rise, heat flux, and contact pressure, which are strictly positive and exhibit varying degrees of right skewness, a log1p transformation is applied to perform nonlinear compression of the data, effectively mitigating the right-skewed distribution and reducing the disproportionate influence of extreme values on model training. Robust scaling is further applied to enhance the stability of the feature representations.

3.2 Model training and hyperparameter optimization

Following dataset generation and normalization, the cU-Net model is constructed and trained to predict the physical field distributions, with the model architecture and theoretical details provided in Section 2.3. Prior to formal training, Bayesian optimization is applied to systematically tune the model hyperparameters[50, 51], ensuring that the model architecture and training strategy are well-adapted to the statistical characteristics of the dataset. A subset of 5,000 samples is randomly drawn from the complete dataset for hyperparameter search, over which 100 independent trials are conducted. The optimal hyperparameter configuration is determined based on the lowest validation loss, and the resulting model architecture and training hyperparameters are summarized in Table 3.

Table 3 Model architecture and training hyperparameters.

Parameter	Value	Parameter	Value
Encoder base channel number	64	Learning rate	9.55×10^{-5}
Encoder-decoder depth	4	Batch size	32
Bottleneck channel number	512	Weight decay	6.63×10^{-7}
Condition encoder hidden dimension	256	MAE loss weight	0.50
Multi-scale attention kernel sizes	1×1, 3×3, 5×5	Gradient loss weight	0.25
Channel attention compression ratio	8	Focal loss weight	0.15
Dropout rate	7.26×10^{-3}	Spectral loss weight	0.10
Learning rate decay factor	0.5	Focal focusing parameter	2.0

Note: The loss weights in the table are the final values obtained in Stage 3 of progressive training.

To evaluate hyperparameter sensitivity, all Bayesian optimization trials were further analyzed using Spearman rank correlation and functional analysis of variance (fANOVA), as shown in Fig. 6. Spearman correlation evaluates the monotonic relationship between each hyperparameter and the validation loss, whereas fANOVA quantifies the nonlinear contribution of each hyperparameter to the overall performance variance.

The results indicate that the learning rate ($\rho = -0.2354$), weight decay ($\rho = 0.2537$), and dropout rate($\rho = 0.2216$) exhibit weak monotonic effects on validation performance ($|\rho| < 0.26$), indicating that model performance is relatively insensitive to moderate variations of these hyperparameters within the explored search ranges. The high-value boost factor shows no clear monotonic relationship but achieves the highest fANOVA importance (0.434), suggesting that its influence is primarily nonlinear and associated with interactions affecting high-value contact regions. By contrast, the focal focusing parameter γ shows negligible influence according to both Spearman correlation ($\rho = -0.0046$) and fANOVA (importance=0.022). Overall, these results indicate that the optimized hyperparameter configuration lies within a relatively flat performance region, demonstrating robustness of the proposed cU-Net to moderate hyperparameter perturbations.

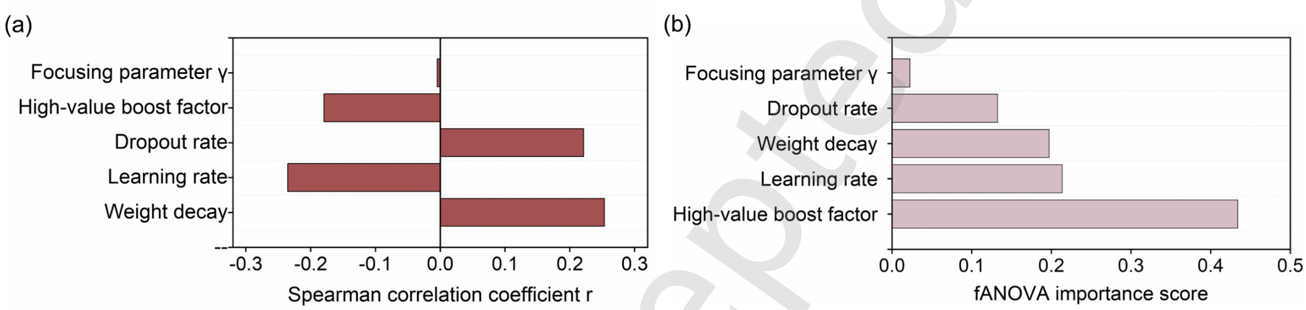

Fig. 6 Hyperparameter sensitivity analysis (a) Spearman rank correlation coefficients (b) fANOVA importance scores.

To quantitatively evaluate prediction performance, complementary error metrics were adopted. The mean relative error (MRE) evaluates the overall prediction accuracy over the entire effective contact region. The peak relative error (Peak-RE) measures the relative error of the maximum response, reflecting the accuracy of extreme-value prediction. The top 5% mean relative error (Top5%-MRE) evaluates the average relative error within the top 5% highest-value regions, providing an assessment of the model's capability to reconstruct the localized extreme values associated with pressure and thermal hotspots. In addition, R^2 is the coefficient of determination computed over all nodes of the effective contact region.

To verify the effectiveness of the proposed progressive composite loss, an ablation study was conducted under identical dataset partitioning, network architecture, optimization strategy, and hyperparameter settings. The only difference among the four groups was the loss formulation. Group A employed MAE only. Group B employed MAE + Gradient loss. Group C employed MAE +

Gradient loss + Focal loss. Group D employed the complete MAE + Gradient loss + Focal loss + Spectral loss, which is the final loss configuration proposed in this work. The quantitative comparison is summarized in Table 4.

The ablation results demonstrate that each loss component contributes to a different aspect of prediction performance. The gradient loss primarily improves the reconstruction of regions with large spatial gradients, resulting in lower Peak-RE. The focal loss further enhances the prediction of sparse localized high-value regions, while the spectral loss improves the consistency of spatial-frequency information between the predicted and reference fields, leading to higher overall reconstruction fidelity. Consequently, the complete four-component loss (Group D) provides the best overall balance among global prediction accuracy, extreme-value prediction, reconstruction of localized high-value regions, and spatial-frequency consistency across the three predicted physical fields, and is therefore adopted as the final training objective.

Table 4 Ablation study of the progressive composite loss. Group A: MAE; Group B: MAE + Gradient; Group C: MAE + Gradient + Focal; Group D: MAE + Gradient + Focal + Spectral.

Group	Temperature			Pressure			Heat flux		
	MRE (%)	Peak-RE (%)	Top5%-MRE (%)	MRE (%)	Peak-RE (%)	Top5%-MRE (%)	MRE (%)	Peak-RE (%)	Top5%-MRE (%)
A	3.58	3.99	3.40	13.18	15.92	3.38	13.92	10.12	4.78
B	3.67	3.55	3.13	12.71	9.70	**3.06**	13.83	9.37	4.35
C	3.60	3.67	3.34	12.40	7.63	3.22	**13.62**	9.14	4.43
D	**3.53**	**3.49**	**2.94**	**12.32**	**7.22**	3.12	13.66	**8.62**	**4.25**

Model training is conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU using PyTorch as the deep learning framework, with the AdamW optimizer employed for parameter updates. A three-stage progressive training strategy was adopted over 500 epochs, gradually introducing the Gradient, Focal, and Spectral losses while dynamically adjusting their weights. The final loss configuration is listed in Table 3. Early stopping (patience = 20 epochs) and an adaptive learning-rate scheduler (decay factor = 0.5 after 10 epochs without validation improvement) were employed to improve convergence stability and generalization.

4 Results and Discussion

4.1 Architecture comparison and prediction accuracy

To verify the necessity of the proposed cU-Net architecture, this section compares cU-Net against two baseline models: an MLP and a Transformer. All three models were trained using identical data splits, loss function, and training strategy, with the network architecture as the sole variable. Evaluation metrics include MRE, Top5%-MRE, and R^2 (as defined in Section 3.2). The comparison results are summarized in Table 5.

As shown in Table 5, the MLP exhibited the poorest performance for all three physical fields. Since the two-dimensional rough-surface height map must be flattened into a one-dimensional vector before entering the fully connected layers, the spatial neighborhood relationship of asperities is completely destroyed. Consequently, the model cannot effectively reconstruct localized contact responses, resulting in large prediction errors for pressure and heat-flux fields (MREs of 69.86% and 74.09%, respectively), with negative coefficients of determination, indicating that the model fails to capture the underlying physical mapping.

The Transformer achieved better performance than the MLP for the temperature field (MRE: 20.68%), but its prediction accuracy deteriorated substantially for the pressure and heat-flux fields. The corresponding MREs exceeded 68%, while the coefficients of determination remained close to zero. This behavior suggests that, under the present dataset scale, the global self-attention mechanism is insufficient for learning the highly localized contact responses dominated by neighboring asperity interactions. Compared with natural-image tasks, the contact pressure and heat-flux distributions contain sparse hotspots and sharp spatial gradients, which are difficult to reconstruct solely through global attention.

Table 5 Performance comparison between the proposed cU-Net and representative baseline architectures.

Model	Params (M)	Temperature			Pressure			Heat flux		
		MRE (%)	Top5% MRE(%)	R^2	MRE (%)	Top5% MRE(%)	R^2	MRE (%)	Top5% MRE(%)	R^2
MLP	46.24	51.44	51.98	−0.5163	69.86	80.84	−0.0657	74.09	80.65	−0.1713
Transformer	26.17	20.68	29.58	0.4866	68.41	66.68	0.1062	78.05	67.25	−0.2775
cUNet	**21.15**	**3.53**	**2.94**	**0.9893**	**12.32**	**3.12**	**0.9849**	**13.66**	**4.25**	**0.9779**

In contrast, the proposed cU-Net consistently achieved the highest prediction accuracy for all three physical fields while using the smallest number of trainable parameters. The temperature, pressure, and heat-flux fields reached MREs of 3.53%, 12.32%, and 13.66%, respectively, with coefficients of determination exceeding 0.97 for all outputs. The superior performance can be attributed to the close correspondence between the network architecture and the physical characteristics of rough-surface contact fields. The convolution-based encoder preserves local spatial neighborhoods and naturally captures asperity-scale interactions, which dominate the formation of contact pressure, flash temperature, and heat-flux distributions. The encoder–decoder architecture with skip connections further recovers fine spatial details that may be weakened during feature compression, enabling accurate reconstruction of contact boundaries and localized hotspots. Moreover, the multi-scale spatial attention modules adaptively enhance discriminative features at different spatial scales, allowing the network to simultaneously focus on point-like peaks, local contact patches, and broader contact regions. Together, these architectural characteristics enable cU-Net to achieve more accurate full-field prediction than both the vector-based MLP and the Transformer baseline.

4.2 Full-field prediction performance

Following the training process described above, the model converges to its optimal state on the validation set and the best checkpoint is saved. Model performance is systematically evaluated from two perspectives: overall prediction accuracy and visual field comparison.\

In practical contacts, a large number of nodes in the computational domain are in non-contact or marginal contact states, with corresponding physical field values approaching zero. Including these near-zero regions in the error statistics would cause singularities in the relative error due to near-zero denominators, compromising the validity of the evaluation. An adaptive threshold masking strategy is therefore employed to identify the effective contact region. The threshold is determined as a weighted sum of the 10th percentile, 25th percentile, and 0.2 times the standard deviation of the absolute values of the target physical field, with weights of 0.3, 0.4, and 0.3, respectively, confining the evaluation to physically meaningful contact regions.

The quantitative prediction errors for the three physical fields on the test set are summarized in Table 6. The MREs for the temperature and pressure fields are 3.53% and 12.32%, respectively. Although the heat flux field exhibits the highest error metrics among the three fields, an MRE of 13.66% still indicates that the model captures the overall spatial trend of the heat flux distribution reasonably well. For all three fields the Peak-RE and the Top5%-MRE are lower than the corresponding global MRE, which shows that the localized extreme values and hotspot regions relevant to failure are reconstructed with even higher accuracy than the field as a whole.

Table 6 Overall prediction errors of the model.

Physical field	MRE	Peak-RE	Top5%-MRE
Temperature	3.53%	3.49	2.94
Pressure	12.32%	7.22	3.12
Heat flux	13.66%	8.62	4.25

Fig. 7 presents a comparison of the input rough surface topography, target physical fields, predicted physical fields, and pixel-wise relative error maps for six representative test samples. For the temperature field, the predicted and target fields show high consistency in overall morphology, high-temperature zone locations, and magnitude range, with the relative error below 2.50% for the vast majority of nodes. Elevated errors are sparsely distributed in edge transition regions with steep temperature gradients, indicating accurate overall temperature field prediction but limited reconstruction capability at locally sharp gradients. For the pressure field, the spatial patterns of high-pressure contact zones and low-pressure non-contact zones are well reproduced, with uniformly low errors inside the contact zone, confirming accurate prediction of the continuous pressure distribution within the contact region. Elevated errors are concentrated at contact boundaries with steep pressure gradients, suggesting limited resolution of sharp boundary transitions. For the heat flux field, the predicted and target fields show good overall agreement in spatial texture and magnitude, with localized high-error points scattered in regions of rapid heat flux variation.

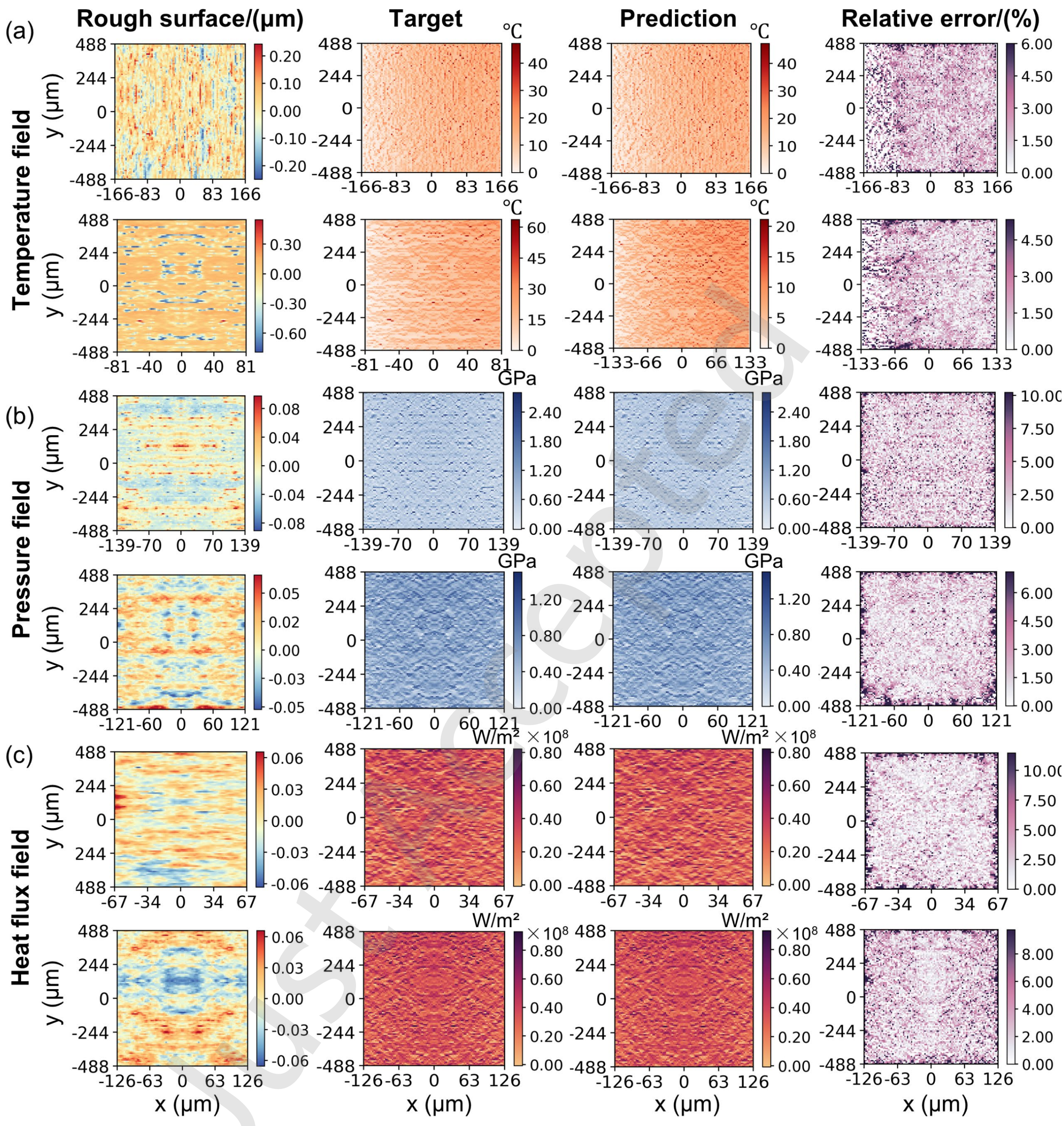

Fig. 7 Predicted physical field distributions showing input rough surface topography, target field, predicted field and relative error distribution for (a) Temperature field (b) Pressure field (c) Heat flux field.

Fig. 8a shows scatter plots of predicted versus ground-truth values for the three physical fields. The data points are tightly clustered along the ideal diagonal in all three cases, indicating strong overall linear correlation and the absence of significant systematic bias. However, data points at extreme high values consistently fall below the diagonal, revealing a persistent underestimation bias. This behavior is attributed to the highly localized spatial nature of extreme field values at asperity contact hot spots, which are difficult for the network to fully capture within its limited receptive field. The scarcity of such extreme samples in the training set further causes the model to regress toward the training mean, with this effect being most pronounced for the heat flux field.

Fig. 8b and c present the kernel density plots and probability density histograms of the relative errors, respectively. The kernel density plots show that the high-density cores of all three physical fields are concentrated in the low-value, low-error region, with the high-density area rapidly compressing toward zero as the absolute field value increases, validating the effectiveness of the adaptive threshold masking strategy. The probability density distributions reveal right-skewed, heavy-tailed error characteristics for all three fields, with the majority of nodes exhibiting low prediction errors and a small number of extreme-value nodes contributing to the long tail. The pressure field shows the largest gap between the mean and median relative errors, indicating the most pronounced heavy-tail behavior, while the temperature field exhibits the smallest gap between the mean and median relative errors, reflecting the most stable and robust prediction performance. In summary, the prediction errors are spatially concentrated in two categories of regions: low-error zones across the majority of the contact area, and localized extreme-value zones at asperity peaks and contact boundaries, with the prediction accuracy of the main contact nodes exceeding the overall mean error level.

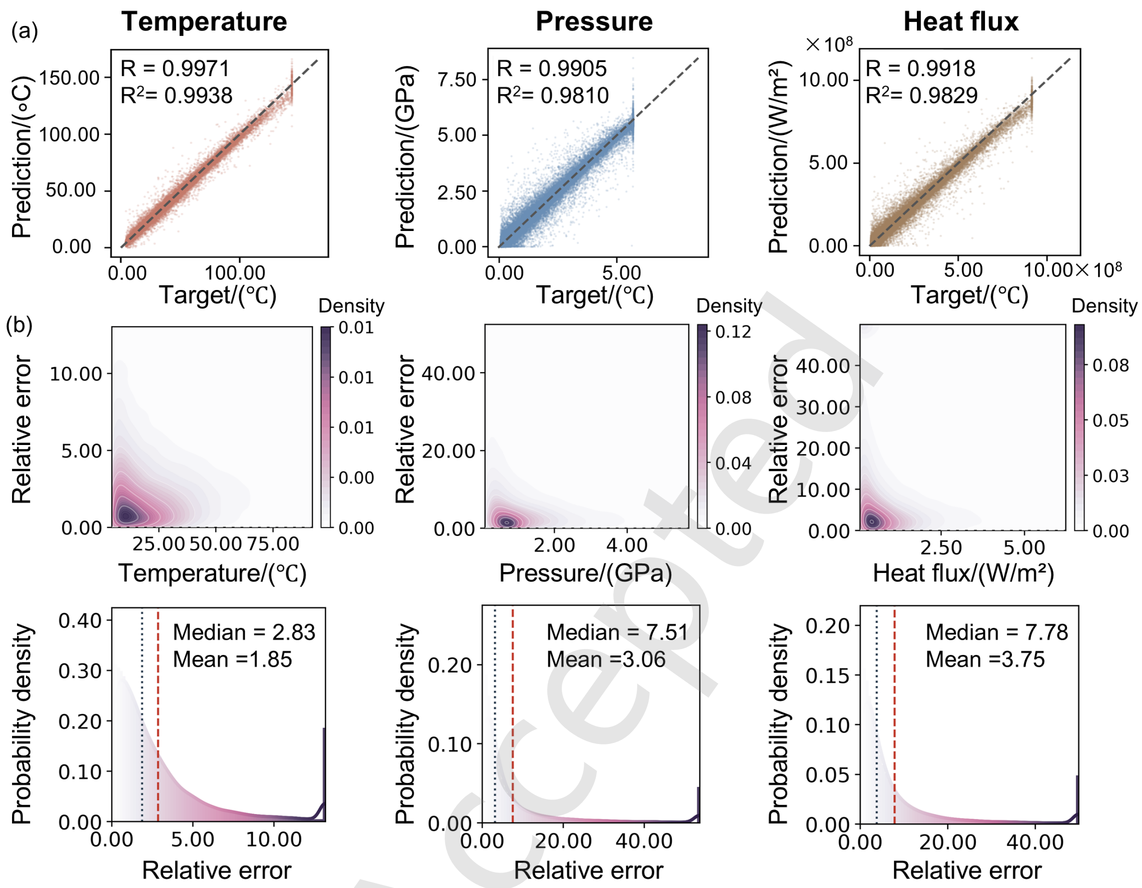

Fig. 8 Statistical analysis of model prediction errors (a) Predicted versus ground-truth scatter plots for different physical fields (b) Kernel density plots of relative errors (c) Probability density histograms of relative errors.

4.3 Operating condition generalization

To quantitatively evaluate the out-of-distribution (OOD) predictive capability of the proposed surrogate model, systematic OOD tests were conducted by independently extrapolating four operating-condition parameters, namely entrainment velocity (velocity), SRR, Load per unit length (load), and equivalent radius (radius). For each parameter, three extrapolation levels were considered, corresponding to approximately 10% (Near), 30% (Mid), and 50% (Far) beyond the training range, resulting in a total of 12 OOD evaluation scenarios.

It should be noted that a surface-isolated partitioning strategy was adopted during dataset construction. Consequently, the measured rough surfaces contained in the test set, together with all their GPNN-generated samples, were completely excluded from the training process. Therefore, the present evaluation simultaneously assesses the predictive capability of the proposed model on previously unseen rough surfaces and its robustness under extrapolated operating conditions. The corresponding MREs of the three predicted physical fields under different OOD scenarios are presented in Fig. 9.

For all four operating parameters, the prediction error generally increases with extrapolation distance. The influence of load is most pronounced on the pressure field, with the MRE increasing from 14.36% (Near) to 37.07% (Far). A similar trend is observed for radius, where the pressure MRE reaches 44.56% at the Far-OOD level, indicating that pressure prediction is highly sensitive to extrapolation of contact geometry and loading conditions.

In contrast, extrapolation of velocity and SRR has only a limited influence on the pressure field, with the MRE remaining approximately 12-13% across all extrapolation levels. This observation is consistent with contact mechanics, where the pressure distribution is primarily governed by load and contact geometry, while being relatively insensitive to kinematic parameters.

For the temperature and heat-flux fields, the prediction error increases progressively with extrapolation distance for all four operating parameters. Compared with the pressure field, these two physical fields exhibit a stronger dependence on operating-condition extrapolation, reflecting their greater sensitivity to variations in frictional heat generation and heat transfer.

Similar to other data-driven surrogate models, the proposed framework is primarily intended for rapid interpolation within the parameter space covered by the training dataset. Although the model demonstrates reasonable robustness under moderate extrapolation, prediction accuracy gradually decreases as the operating conditions move further away from the training domain. Therefore, the proposed surrogate model is recommended for applications within the investigated operating-condition range, whereas predictions beyond this range should be interpreted with appropriate caution according to the OOD evaluation results.

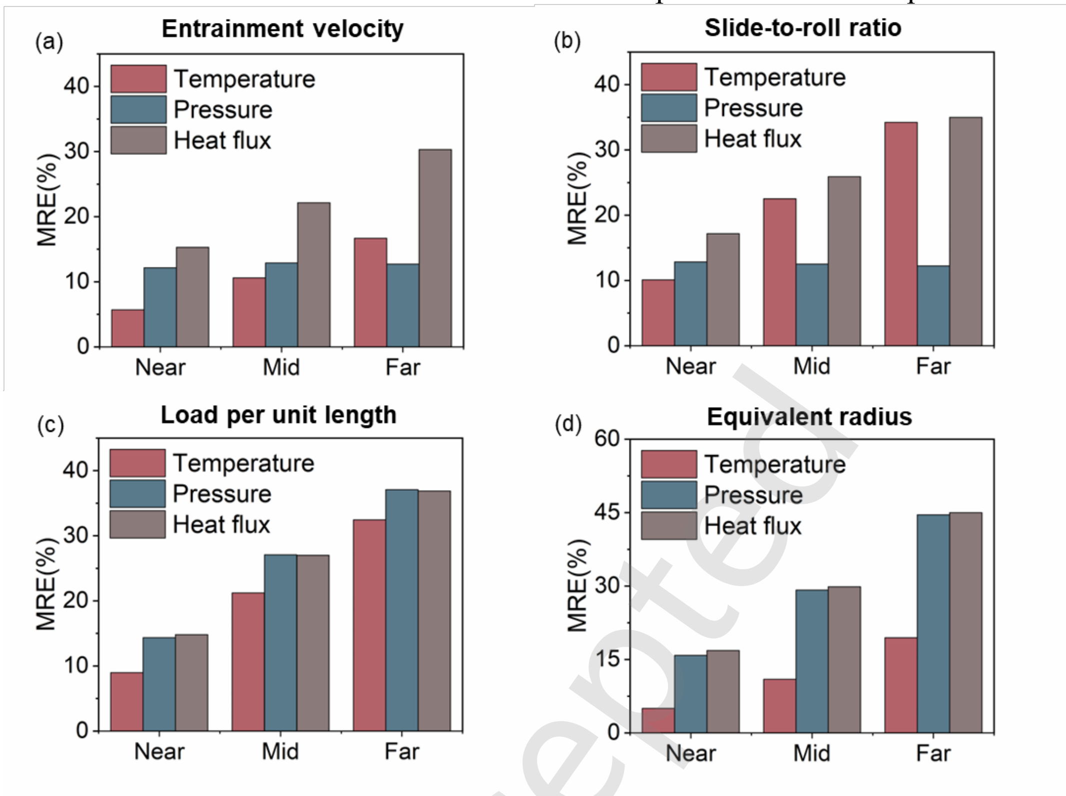

Fig. 9 OOD prediction performance under extrapolated operating conditions (a) Entrainment velocity (b) Slide-to-roll ratio (c) Load per unit length (d) Equivalent radius.

4.4 Computational efficiency and model limitations

Fig. 10 compares the computational cost of the physics-based model and the proposed surrogate model. The physics-based model was evaluated on both a local workstation and a high-performance computing (HPC) platform. The local workstation was equipped with an Intel Core i5-10400F CPU (6 cores, 12 threads, 2.90 GHz), while the HPC platform employed a single compute node with 128 AMD EPYC 7H12 CPU cores (2.60 GHz). The surrogate model was trained and evaluated on the same HPC platform, using an NVIDIA GeForce RTX 4090 GPU. Under boundary lubrication conditions, a single evaluation of the physics-based model required approximately 30 min on the local workstation and 5 min on the HPC platform, whereas the surrogate model completed the prediction in approximately 1 ms. Because the surrogate model was timed on the HPC platform, the speedup is referenced consistently to the HPC implementation of the physics-based model, giving approximately 10^5. The local workstation time is reported only to indicate the cost on commodity hardware.

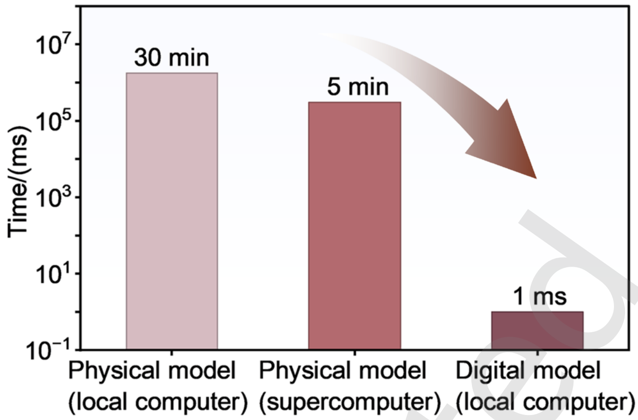

Fig. 10 Computational efficiency comparison between the physics-based model and the digital surrogate model.

Nevertheless, as a data-driven surrogate model, its generalization capability is inherently constrained by the distribution of the training data. Prediction accuracy degrades when the operating conditions fall outside the parameter space covered by the training dataset, as quantified by the out-of-distribution tests of Section 4.3. Prediction reliability may also deteriorate when the statistical characteristics of the target rough surface, such as roughness and height distribution skewness, differ substantially from those of the training samples.

Two further restrictions follow from the way the model is constructed. The framework takes the surface topography as an input and is therefore intended for regimes in which the topography is statistically stable, typically the steady-state wear stage. During running-in or severe wear the topography evolves continuously, and the limiting difficulty is then not only the changing morphology but also the impossibility of acquiring the instantaneous three-dimensional topography in real time. In addition, as discussed in Section 2.3, the trained network is tied to the specific grid resolution used to generate the dataset, so that height maps measured at a different resolution must be resampled before inference and a substantially different discretization requires retraining.

It should be emphasized that the proposed surrogate model is intended to provide rapid prediction of the solutions generated by the underlying physics-based model, rather than to directly replace real friction experiments. Consequently, its applicability is inherently bounded by the assumptions and validity range of the adopted boundary-lubrication model. Nevertheless, the proposed framework is highly extensible. For other lubrication regimes, such as mixed lubrication and EHL, full-field prediction of rough-surface contact parameters can be achieved by regenerating the training dataset using the corresponding physics-based model and retraining the network. Therefore, the present framework provides a general surrogate modeling strategy that can be readily extended to different lubrication conditions while retaining the computational efficiency of deep learning-based prediction.

5 Conclusions

This work proposes a cU-Net-based digital surrogate model for predicting the spatial distributions of contact parameters, including contact pressure, temperature, and heat flux, on rough surfaces in line-contact friction.

Real surface topographies measured using an optical 3D surface profiler are augmented via the GPNN model, and a systematic training dataset covering a wide range of operating conditions and diverse surface topographies is constructed through batch computation using a boundary lubrication physics-based model, providing a reliable data foundation for surrogate model training.

The cU-Net surrogate model is developed with an operating condition embedding mechanism for unified multi-condition prediction, multi-scale spatial and channel attention modules for enhanced cross-scale feature extraction, and a four-component progressive composite loss function tailored to the sparse extreme-value characteristics of rough-surface contact parameters. Combined with Bayesian hyperparameter optimization and a three-stage progressive training strategy, the model effectively constrains the gradient features and frequency-domain consistency of the predicted contact parameter distributions.

Systematic evaluation on completely unseen rough surfaces gives mean relative errors of 3.53%, 12.32%, and 13.66% for the temperature, pressure, and heat flux fields, respectively, with peak and top-5% relative errors that are lower still, so that the localized extreme values relevant to failure are captured at least as accurately as the field as a whole. High-error regions are primarily concentrated at asperity contact edges and at contact boundaries with steep field gradients. The error distribution exhibits a right-skewed heavy-tail characteristic, with the temperature field showing the most stable and robust prediction performance. Out-of-distribution tests quantify how the accuracy degrades once the operating conditions are extrapolated beyond the training envelope.

The surrogate model achieves a computational speedup of approximately 10^5 over the physics-based model, providing a practical foundation for engineering applications such as rapid parameter design, real-time condition monitoring, and digital twin construction. The proposed model framework is not limited to boundary lubrication conditions and, given training datasets generated from appropriate physics-based models, is expected to be applicable to full-field prediction of contact parameters such as contact pressure, temperature, and film thickness under different lubrication regimes.

Nomenclature

Symbol	**Description**
$\mathbf{A}_c$	Channel attention output
$\mathbf{A}_s$	Multi-scale spatial attention output
C	Number of feature channels
C_d	Number of channels at the deepest encoder layer
$\mathbf{c}_{emb}$	Condition embedding vector
$\mathbf{F}$	Input feature map
$\mathbf{F}_{bottleneck}$	Feature map at the network bottleneck
$\mathbf{F}_d^{enc}$	Feature map at the deepest encoder layer
$G_x(\cdot), G_y(\cdot)$	Sobel operators extracting spatial gradients in x and y directions
H	Spatial height of a feature map (pixels)
H_d	Spatial height of the deepest encoder feature map (pixels)
$\mathbf{1}_{H_d \times W_d}$	All-ones matrix of size Hd×WdH_d \times W_d
L_{focal}	Focal loss
L_{grad}	Adaptive gradient loss
L_{MAE}	Mean absolute error loss
$L_{spectral}$	Spectral loss
L_{total}	Total composite loss
m_i	Sample weight mask assigning higher weights to high-value regions

N	Total number of pixels
r	Channel compression ratio in channel attention
u,v	Frequency domain coordinates
W	Spatial width of a feature map (pixels)
W_d	Spatial width of the deepest encoder feature map (pixels)
W_c^1, W_c^2	Trainable dimensionality reduction weight matrix in channel attention
$\hat{y}_i$	Predicted value of the physical quantity at pixel i
y_i	Ground truth value of the physical quantity at pixel i
$\mathcal{F}(\cdot)$	Amplitude spectrum of the 2D fast Fourier transform
$\alpha_{x,i}, \alpha_{y,i}$	Adaptive gradient weights in x and y directions at pixel i
β_k	Softmax-normalized learnable fusion weight in multi-scale spatial attention
γ	Focal loss focusing parameter
λ_i	Weighting coefficient for each loss component
$\sigma(\cdot)$	Sigmoid activation function

Acknowledgements

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

Funding

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared. The names of funding organizations should be written in full.

Author contributions

Yichun Xia: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing–original draft, Visualization.

Hui Cao: Conceptualization, Writing – review and editing.

Yonggang Meng: Conceptualization , Resources, Writing–review and editing, Supervision, Project administration, Funding acquisition.

Availability of data and materials

Not applicable.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Chen T et al. A review on gear scuffing studies: Theories, experiments and design. *Tribology International* 196: 109741 (2024).

[2] Braumann L et al. Tribology technologies for gears in loss of lubrication conditions: a review. *J Mater. Sci: Mater Eng.* 20(1): 29 (2025).

[3] Li M et al. A comprehensive dynamic study of the cylindrical roller bearings utilizing the oil film characteristics and thermal-mechanical coupling mechanisms. *Int J Mech Mater Des* 22(1): 21 (2026).

[4] Liu H et al. Study on contact fatigue of a wind turbine gear pair considering surface roughness. *Friction* 8(3): 553–567.

[5] Jangra D et al. Gear faults diagnosis: a comprehensive review of experimental and theoretical studies. *Structural Health Monitoring* : 14759217251324698 (2025).

[6] Duan Z et al. Analysis of the line contact tribo-lubrication pair and failure mechanism under the extreme conditions. *Tribology International* 185: 108505 (2023).

[7] Lyu B et al. A deterministic contact evolution and scuffing failure analysis considering lubrication deterioration due to temperature rise under heavy loads. *Engineering Failure Analysis* 123: 105276 (2021).

[8] He T et al. Experimental and numerical investigations of the Stribeck curves for lubricated counterformal contacts. *Journal of Tribology* 139(2): 021505 (2017).

[9] Zhu D, Hu Y-Z. The study of transition from elastohydrodynamic to mixed and boundary lubrication. *The advancing frontier of engineering tribology, Proceedings of the 1999 STLE/ASME HS Cheng Tribology Surveillance* : 150–156 (1999).

[10] Kennedy F E et al. Thin film thermocouple arrays for sliding surface temperature measurement. *Wear* 207(1): 46–54 (1997).

[11] Helmig T et al. An infrared thermography based experimental method to quantify multiscale thermal resistances at non-conforming interfaces. *International Journal of Heat and Mass Transfer* 186: 122399 (2022).

[12] Gohar R, Cameron A. Optical Measurement of Oil Film Thickness under Elasto-hydrodynamic Lubrication. *Nature* 200(4905): 458–459 (1963).

[13] Cen H et al. EHL film thickness in rolling element bearings evaluated by electrical capacitance method: A review. *Tribology - Materials, Surfaces & Interfaces* 15(1): 55–77 (2021).

[14] Ebner M et al. Measurement of EHL temperature by thin film sensors – Thermal insulation effects. *Tribology International* 149: 105515 (2020).

[15] Lu J et al. 3D Measurements of Lubricant and Surface Temperatures Within an Elastohydrodynamic Contact. *Tribol Lett* https://doi.org/10.1007/s11249-017-0953-2 (2018).

[16] Hu Y-Z, Zhu D. A Full Numerical Solution to the Mixed Lubrication in Point Contacts. *Journal of Tribology* 122(1): 1–9 (2000).

[17] Habchi W et al. Stabilized fully-coupled finite elements for elastohydrodynamic lubrication problems. *Advances in Engineering Software* 46(1): 4–18 (2012).

[18] Ren N et al. A Three-Dimensional Deterministic Model for Rough Surface Line-Contact EHL Problems. *J. Tribol* https://doi.org/10.1115/1.2991291 (2008).

[19] Liu S et al. A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear* 243(1): 101–111 (2000).

[20] Tian X, Kennedy F E Jr. Contact Surface Temperature Models for Finite Bodies in Dry and Boundary Lubricated Sliding. *J. Tribol* 115(3): 411–418 (1993).

[21] Jaeger J C. Moving sources of heat and the temperature at sliding contacts. *Journal and proceedings of the Royal Society of New South Wales* 76(3): 203–224 (1943).

[22] BLOK H. Theoretical Study of Temperature Rise at Surfaces of Actual Contact under Oiliness Lubricating Conditions. *Proc. Instn. Mech. Engrs. (General discussion on lubrication and lubricants)* 2: 222 (1937).

[23] Marian M et al. Predicting EHL film thickness parameters by machine learning approaches. *Friction* 11(6): 992–1013 (2023).

[24] Issa J et al. Machine learning for film thickness prediction in elastohydrodynamic lubricated elliptical contacts. *Lubricants* 11(12): 497 (2023).

[25] Tošić M et al. Application of machine learning for film thickness prediction in elliptical EHL contact with varying entrainment angle. *Tribology International* 199: 109940 (2024).

[26] Wang L et al. Application of artificial neural networks for predicting friction coefficient and film thickness considering non-Newtonian and thermal effects. *J Supercomput* 81(8): 876 (2025).

[27] Wang Y et al. Development of a machine learning-based surrogate model for friction prediction in textured journal bearings. *Friction* 13(8): 9441051 (2025).

[28] Singh A et al. Machine learning based surrogate modelling for the prediction of maximum contact temperature in EHL line contacts. *Tribology International* 179: 108166 (2023).

[29] Prajapati D K et al. Machine learning approach for the prediction of mixed lubrication parameters for different surface topographies of non-conformal rough contacts. *Industrial lubrication and tribology* 75(9): 1022–1030 (2023).

[30] Kalliorinne K et al. Artificial Neural Network Architecture for Prediction of Contact Mechanical Response. *Front. Mech. Eng.* https://doi.org/10.3389/fmech.2020.579825 (2021).

[31] Suman S, Prajapati D K. Predictive Modeling of Real Contact Area on Rough Surfaces Using Deep Artificial Neural Network. *J. Tribol* https://doi.org/10.1115/1.4068057 (2025).

[32] Hess N, Shang L. Development of a machine learning model for elastohydrodynamic pressure prediction in journal bearings. *Journal of Tribology* 144(8): 081603 (2022).

[33] Kelley J et al. Enhancing practical modeling: a neural network approach for locally-resolved prediction of elastohydrodynamic line contacts. *Tribology International* 199: 109988 (2024).

[34] Cartwright S et al. A machine learning-driven approach to predicting thermo-elasto-hydrodynamic lubrication in journal bearings. *Tribology International* 196: 109670 (2024).

[35] Wu J et al. The SA-UNet temperature field reconstruction based on GAN data augmentation of wet friction components. *Measurement* 254: 117877 (2025).

[36] Jang I, Jang Y H. PREDICTION OF CONTACT DISTRIBUTION ON ROUGH SURFACES USING DEEP LEARNING ALGORITHMS. *Facta Universitatis, Series: Mechanical Engineering* 23(4): 757–785 (2025).

[37] Xia Y et al. Multiscale Simulation and Analysis of Contact Temperature in Line Contacts Under Boundary Lubrication. *Journal of Tribology* https://doi.org/10.1115/1.4070695 (2026).

[38] Majumdar A, Tien C L. Fractal characterization and simulation of rough surfaces. *Wear* 136(2): 313–327 (1990).

[39] Pérez-Ràfols F, Almqvist A. Generating randomly rough surfaces with given height probability distribution and power spectrum. *Tribology International* 131: 591–604 (2019).

[40] Granot N et al. Drop the gan: In defense of patches nearest neighbors as single image generative models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022: 13460–13469.

[41] Ronneberger O et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015: 234–241.

[42] Hu J et al. Squeeze-and-Excitation Networks//Proceedings. In: *the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.

[43] Szegedy C et al. Going Deeper with Convolutions//Proceedings. In: *the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1–9.

[44] Li X et al. Selective kernel networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019: 510–519.

[45] Lin T-Y et al. Focal loss for dense object detection. In: 2017: 2980–2988.

[46] Fukami K et al. Super-Resolution Reconstruction of Turbulent Flows with Machine Learning. *Journal of Fluid Mechanics* 870: 106–120 (2019).

[47] Liu B et al. Deep Learning Methods for Super-Resolution Reconstruction of Turbulent Flows. *Physics of Fluids* 32(2): 025105 (2020).

[48] Jiang L et al. Focal Frequency Loss for Image Reconstruction and Synthesis. In: 2021: 13919–13929.

[49] Cheng R et al. Improved super-resolution reconstruction of turbulent flows with spectral loss function. *Physics of Fluids* 37(3): 035208 (2025).

[50] Falkner S et al. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. In: *Proceedings of the 35th International Conference on Machine Learning*, 2018: 1437–1446.

[51] Turner R et al. Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020//NeurIPS 2020 Competition and Demonstration Track. In: *PMLR*, 2021: 3–26.

Yichun Xia received his B.S. degree from Dalian University of Technology, Dalian, China, in 2019, and his M.S. degree from the State Key Laboratory of Tribology in Advanced Equipment (SKLT), Tsinghua University, Beijing, China, in 2022, where he is currently pursuing his Ph.D. degree. His research interests include elastohydrodynamic lubrication modeling and friction temperature prediction in tribological contacts.

Dr.Yonggang Meng, tenured full Professor in Mechanical Engineering, Tsinghua University, and Adjunct Professor in the State Key Laboratory of Tribology in Advanced Equipment. His research area covers over lubrication theory and its applications, friction mechanisms and active control technology, wear of materials and its control as well as tribological design of machine components. He is the winner of National Award of Technology Invention, National Award of Progress in Science and Technology, and Natural Science Award of MOE. He is the author or co-author of more than 340 peer-reviewed publications, and the holder of more than 20 China and/or PCT patents.