

<https://doi.org/10.1038/s44172-025-00508-0>

# A deep learning framework for predicting the effect of surface topography on thermal contact resistance

Man Zhou<sup>✉</sup>, Zhuoyan He, Peiyao Guo, Ping Zhang<sup>✉</sup>, Lin Tao<sup>✉</sup> & Mengjun Chen

Efficient heat transfer across contacting surfaces is essential for effective thermal management; however, it is often hindered by thermal contact resistance resulting from complex surface topography. Here we present an interpretability analysis utilizing deep learning based to predict and visualize the key features influencing thermal contact resistance. We developed a convolutional-neural-network-based model trained on an extensive dataset generated using surface fractal theory. The model's predictive performance was validated against experimental data, in which surface topography and thermal resistance were measured for ground and turned steel specimens. Our model accurately predicts thermal contact resistance and estimates the actual contact area. Moreover, interpretability and visualization techniques reveal that both direct contact spots and non-contact regions of the surface topography significantly affect heat transfer, surpassing the explanatory power of traditional roughness parameters. This approach provides a robust methodology to enhance the fundamental understanding and predictive capabilities regarding thermal contact resistance.

The state-of-the-art development of technologies such as high-performance computing<sup>1</sup>, next-generation chips<sup>2</sup>, and renewable energies<sup>3</sup> generates substantial heat densities. Thermal contact resistance (TCR) plays a crucial role in applications requiring precise thermal management<sup>4</sup>. Even minimal TCR at interfaces can strongly degrade computational performance or destabilize qubits<sup>5</sup>. Addressing TCR not only resolves immediate thermal bottlenecks but also facilitates breakthroughs in sustainability, miniaturization, and technologies for extreme environments, thereby establishing TCR as a fundamental aspect of modern thermal engineering<sup>6</sup>.

Surface characteristics have long been recognized as critical factors influencing the TCR<sup>6</sup>. Extensive experimental research has demonstrated that surface roughness markedly affects TCR<sup>7–11</sup>. However, Cui et al.<sup>12</sup> highlighted that even when surfaces have identical roughness values, variations in processing methods can result in substantial differences in TCR. Statistical parameters provide an alternative method for characterizing surfaces. Several studies have employed surface fractal models to analyze contact phenomena. Ma et al. conducted experimental measurements of surface fractal parameters, considering elastic, plastic, and elastoplastic deformations, and developed a theoretical prediction model for TCR<sup>13</sup>. Similarly, Sun et al. extended this approach to cylindrical contact surfaces<sup>14</sup>. Chen et al. further enhanced the analysis by incorporating thermal stress and asperity interactions<sup>15</sup>. In summary, these studies characterize surface properties either by using surface roughness metrics, such as arithmetic

average roughness (Ra) and root mean square roughness (Rq), or by deriving fractal parameters, including fractal dimension and scaling factor, from fractal theory. However, theoretical models incorporating complete surface topography have not yet been explored.

An alternative research approach involves directly constructing surface morphology using finite element method (FEM). Dai et al. reconstructed the contact surface and performed FEM simulations with ABAQUS to investigate the TCR of Ti-6Al-4V<sup>16</sup>. They further incorporated thermal expansion effects and conducted thermo-mechanical coupling simulations<sup>17</sup>. Wang et al. studied the TCR under non-uniform loading conditions<sup>18</sup>, while Dong et al. examined the influence of gap conduction<sup>19</sup>. Although FEM-based numerical simulations can accurately represent complete surface morphology, they have considerable drawbacks, including high computational resource demands, long computation times, and limited reusability. These limitations hinder the development of simple and user-friendly general models.

In recent years, neural network-based predictive models have gained increasing attention in researches<sup>20</sup>. Ren et al. developed a predictive model for TCR utilizing artificial neural networks (ANN), enabling accurate determination of TCR values directly from temperature distributions and contact pressure at multiple points, thus eliminating the need for complex experimental testing procedures<sup>21</sup>. Moreover, Feng et al. employed ANN to forecast the TCR of copper blocks subjected to load cycling<sup>22</sup>. However,

School of Mechanical and Electrical Engineering, Guilin University of Electronic technology, Guilin, Guangxi, China. ✉e-mail: pingzhang@guet.edu.cn; Taolin@guet.edu.cn

current investigations on TCR using neural networks primarily rely on simple ANN architectures, overlooking the potential benefits of more sophisticated network designs and their intrinsic properties.

In this study, we develop and validate a deep learning model based on convolutional neural networks (CNNs) that directly utilizes complete surface topography data to predict TCR. The model exhibits robust predictive performance for both TCR and real contact area when evaluated on a comprehensive synthetic dataset derived from a theoretical framework. Subsequently, the model is validated using experimental measurements from ground and turned steel surfaces. By employing interpretability and visualization techniques, we provide direct evidence identifying the specific surface characteristics that influence heat transfer. Our findings reveal that both contact points and, notably, non-contact regions of interacting surfaces strongly affect prediction outcomes. This approach offers a methodology for advancing the fundamental understanding of TCR.

## Results

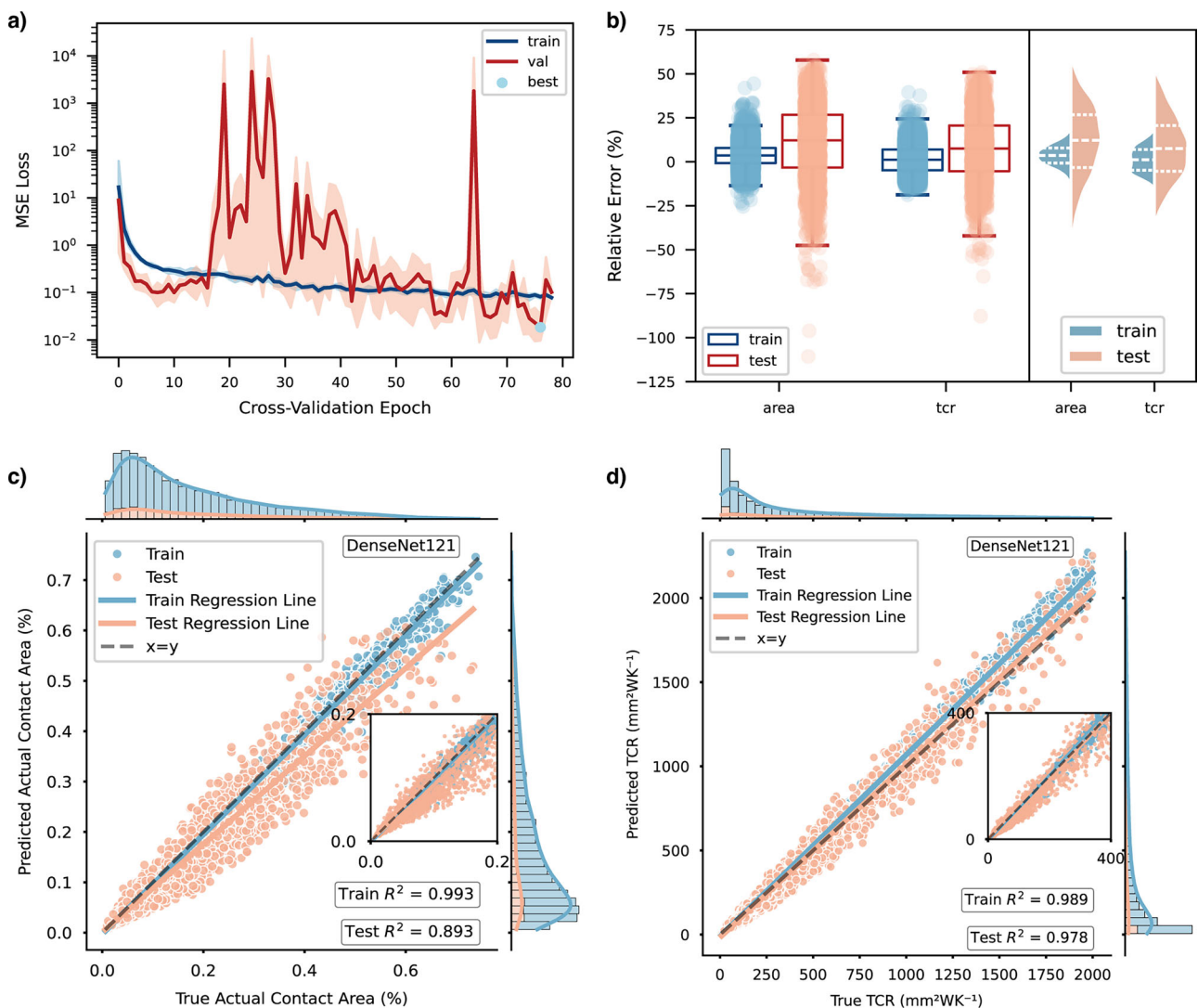
### Model performance

As illustrated in Fig. 1a, the training process was conducted over 80 epochs, with the optimal model identified at the 76th epoch, achieving an average validation mean squared error loss of 0.01 during cross-validation. The

model's predictive performance was evaluated on both the training and validation datasets. The results indicate that the predicted actual contact area achieved determination coefficients ( $R^2$ ) of 0.993 on the training set and 0.893 on the test set (Fig. 1c). Similarly, the TCR yielded  $R^2$  values of 0.989 and 0.978 on the training and test sets, respectively (Fig. 1d). Relative errors were also calculated, as shown in Fig. 1b. For both prediction targets, relative errors were predominantly below 25% on the training set and below 50% on the test set, with maximum relative errors not exceeding 50 and 125%, respectively, indicating robust predictive accuracy. The regression lines for the actual contact area on both datasets lay below the  $x = y$  line, suggesting a tendency of the model to underestimate the actual contact area. In contrast, the regression lines for the TCR were positioned above the  $x = y$  line, implying an overestimation of the TCR.

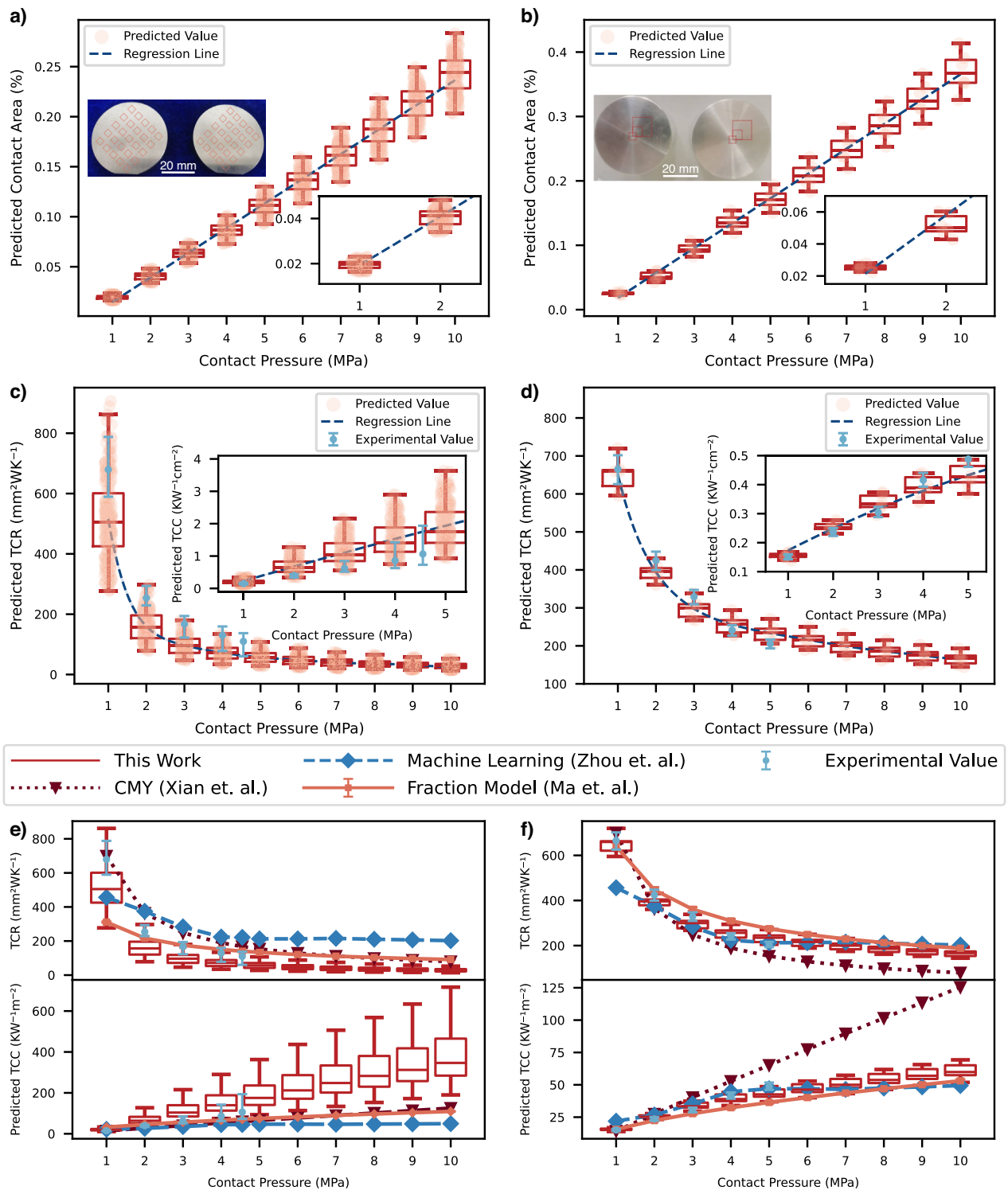
### Experimental validation

After validating the model's performance, further verification was conducted using experimentally measured data. The test specimens are divided into two groups: the first group was fabricated through grinding, as illustrated in Fig. 2a, while the second group was produced by turning (Fig. 2b). Both groups exhibit an Ra of approximately  $0.8 \mu\text{m}$ . Based on processing characteristics, a  $5 \times 5$  grid of local areas was selected for



**Fig. 1 | Training and evaluation details of deep learning model.** a Changes in the mean squared error (MSE) loss during training and validation across the five-fold cross-validation process. The shaded area represents the range of the loss distribution for each epoch, whereas the solid line indicates the corresponding average loss.

b Distribution of relative errors of the prediction model on both the training and test sets. c Performance of the prediction model in predicting the actual contact area. d Performance of the prediction model in predicting thermal contact resistance (TCR).



**Fig. 2 | Comparisons between prediction results, experimental data, and other prediction models. a, b** The predicted actual contact area for **a** ground and **b** turned surfaces. **c, d** The predicted thermal contact resistance (TCR) aligned with measured values for **c** ground and **d** turned surfaces. **e, f** Predicted TCR, and thermal contact

conductance (TCC) compared with the Cooper-Mikic-Yovanovich (CMY) model, machine learning model, and fraction model for **e** ground and **f** turned surfaces. Specimens fabricated using grinding and turning are shown in **(a)** and **(b)**, respectively.

surface topography testing in the first group, whereas three local areas were tested from the center outward in the second group. TCR was measured under pressures ranging from 1 to 4.55 MPa. Surface topography data and contact pressure were utilized as model inputs to predict TCR. Test areas of upper and lower specimens were paired one-to-one, and the predicted data were employed to assess TCR at specific pressures.

Consequently, the first group yielded 625 data sets per pressure, while the second group produced nine data sets.

The experimental data and prediction results are presented in Fig. 2a–d. The predicted actual contact area exhibited a nearly linear increase with pressure, rising from approximately 0.02% to 0.25% and 0.4% as pressure increased from 1 to 10 MPa. Although Ra was controlled to be

similar, turned surfaces tended to have sharper peaks that deform differently under pressure compared to ground surfaces, which had more rounded asperities. At low pressures, both surfaces primarily underwent plastic deformation, resulting in similarly small contact areas proportional to material hardness. However, as pressure increased, elastic deformation became dominant, causing different responses. The taller asperities on the turned surface deformed elastically over a larger area, resulting in a greater actual contact area overall. This phenomenon is detailed in Supplementary Discussion 3. Regarding TCR, model predictions were slightly lower for pressures below 1 MPa and slightly higher for pressures above 3 MPa. Nevertheless, all predictions remained within the acceptable error range. Supplementary Figs. 1–2 present specific surface prediction results. Pronounced differences in local morphological characteristics between surface pairs led to clearly divergent predictive outcomes.

### Comparison with existing models

As shown in e, f, comparative analyses were performed between the predictions of this study and those from the correctional Cooper-Mikic-Yovanovich (CMY) model<sup>23</sup>, the fractal model, and the machine learning (ML) model<sup>24</sup>. For the first set of specimens, experimental measurements aligned more closely with the fractal model predictions within the 2–4.55 MPa range. The ML model exhibited substantial deviations across all pressure levels. Notably, the predictions from this study were consistently lower than those of the other models for this specimen group. For the second set of specimens, the predictions showed closer agreement with the fractal model in the 1–3 MPa range and better alignment with ML predictions at 2–5 MPa. The CMY model failed to predict accurately when the contact pressure exceeded 2 MPa. It is important to note that both the CMY and ML models do not utilize complete surface information, resulting in identical predictions for both specimen groups. In contrast, the fractal model and our proposed model effectively incorporated comprehensive surface characteristics, thereby demonstrating prediction specificity for different surfaces. Detailed descriptions of the CMY and ML models are provided in Supplementary Method 4, along with the extraction methodology of surface characterization parameters for the fractal model in Supplementary Fig. 5 and Supplementary Methods 3.

### Interpretability and visualization

To enhance the interpretability of the model, we employ guided backpropagation (GBP)<sup>25</sup> and Class Activation Mapping (CAM)<sup>26</sup> to visualize the surfaces influencing the final outcomes. Figure 3a, b, and Fig. 3g, h present GBP results for two sets of surfaces. For improved visualization clarity, the surfaces are manually rotated by 90°, 180°, and 270°, preserving the integrity of the input data. The gradients obtained from GBP are further processed to produce output gradients matching the dimensions of the input surfaces. These gradients are divided into grad1 and grad2 along the first dimension, corresponding to gradients of different surfaces. Positive and negative influences are defined as the absolute values of gradients greater than or less than zero, respectively. The subscripts 1 and 2 denote their respective gradients. For the first set of specimens, rotation substantially influences the actual contact area (Fig. 3c) and the corresponding TCR (Fig. 3d) at a contact pressure of 1 MPa. Specifically, an increase in the actual contact area results in a reduction in TCR. However, at contact pressures of 2 MPa or higher, the effect of rotation becomes negligible. For the second set of specimens, rotation has negligible effects on both the TCR (Fig. 3j) and the actual contact area (Fig. 3i), likely due to the manufacturing process. The rotational symmetry of the texture formed by turning along the axis rendered rotation inconsequential. To further analyze the gradients, we summed all their elements, revealing a positive correlation between the sum of grad elements and the actual contact area, as shown in Fig. 3e, and a negative correlation with the TCR in Fig. 3f. In contrast, the

second set of specimens exhibited no discernible trend (Fig. 3k, l), potentially due to minimal variations in the actual contact area and TCR. Examination of the gradient activation maps revealed that they predominantly corresponded to regions of higher or lower surface contact areas (i.e., surf1 + surf2), consistent with TCR theory<sup>6</sup>. This consistency confirms that our model accurately captures the relevant features of the real surfaces.

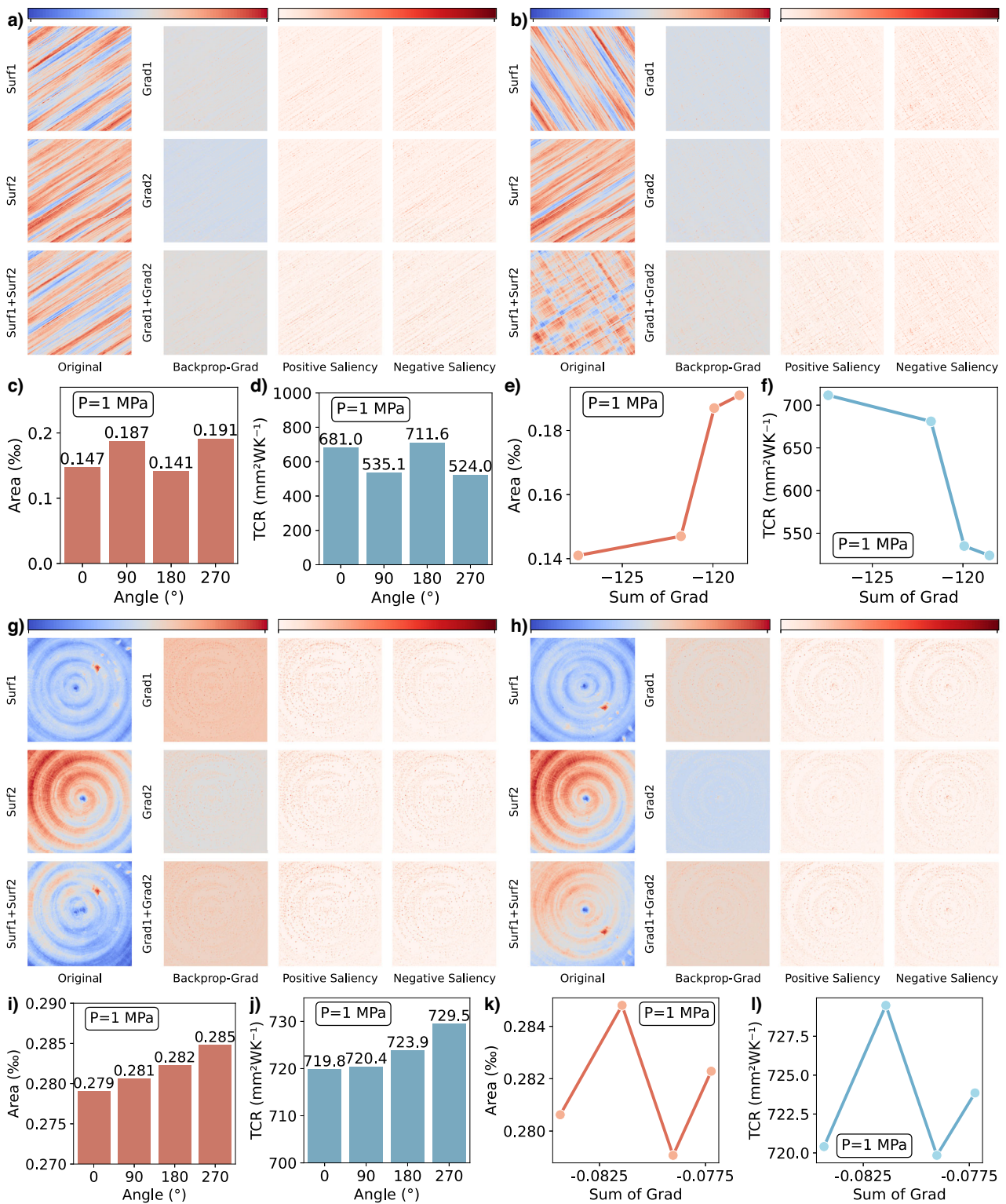
The CAM method demonstrates distinct characteristics in Fig. 4. Most notably, for surfaces, the CAM visualization results emphasize the lower regions of the contact surface. Specifically, the activation maps corresponding to the lower regions of surf1 and surf2 exhibit higher values and exert a greater influence on the outcomes. These lower regions can be interpreted as areas where contact is less likely to occur. Moreover, non-contact areas clearly impact the actual contact area: the greater the extent of non-contact regions, the smaller the actual contact area, resulting in an increased TCR. For all CAM methods, activation maps of shallow layers are more interpretable to humans, whereas those from deeper layers highlight regions that are more difficult to interpret, often showing a stronger emphasis on certain edge regions. We propose two explanations for this phenomenon. First, features extracted by deeper neural networks are inherently more complex and less intuitive. Second, due to the CAM generation process, deeper layer features typically have smaller spatial dimensions but more channels, which can lead to information loss or the introduction of irrelevant information when these features are upsampled to match the input size. Consequently, interpreting activation maps from deeper neural network layers remains a challenging area warranting further investigation.

### Discussion

In summary, this study effectively leverages the robust feature extraction capabilities of CNNs to develop a TCR prediction model based on complete surface morphology data. The model also provides a reference for the actual contact area between two surfaces. As shown in Supplementary Table 1, our model achieves a four-order-of-magnitude acceleration in prediction speed compared to conventional FEM approaches under similar computational configurations. This substantial performance improvement remains evident even when accounting for training overhead, highlighting a notable computational advantage. Moreover, the visualization techniques employed effectively illustrate the surface features influencing TCR.

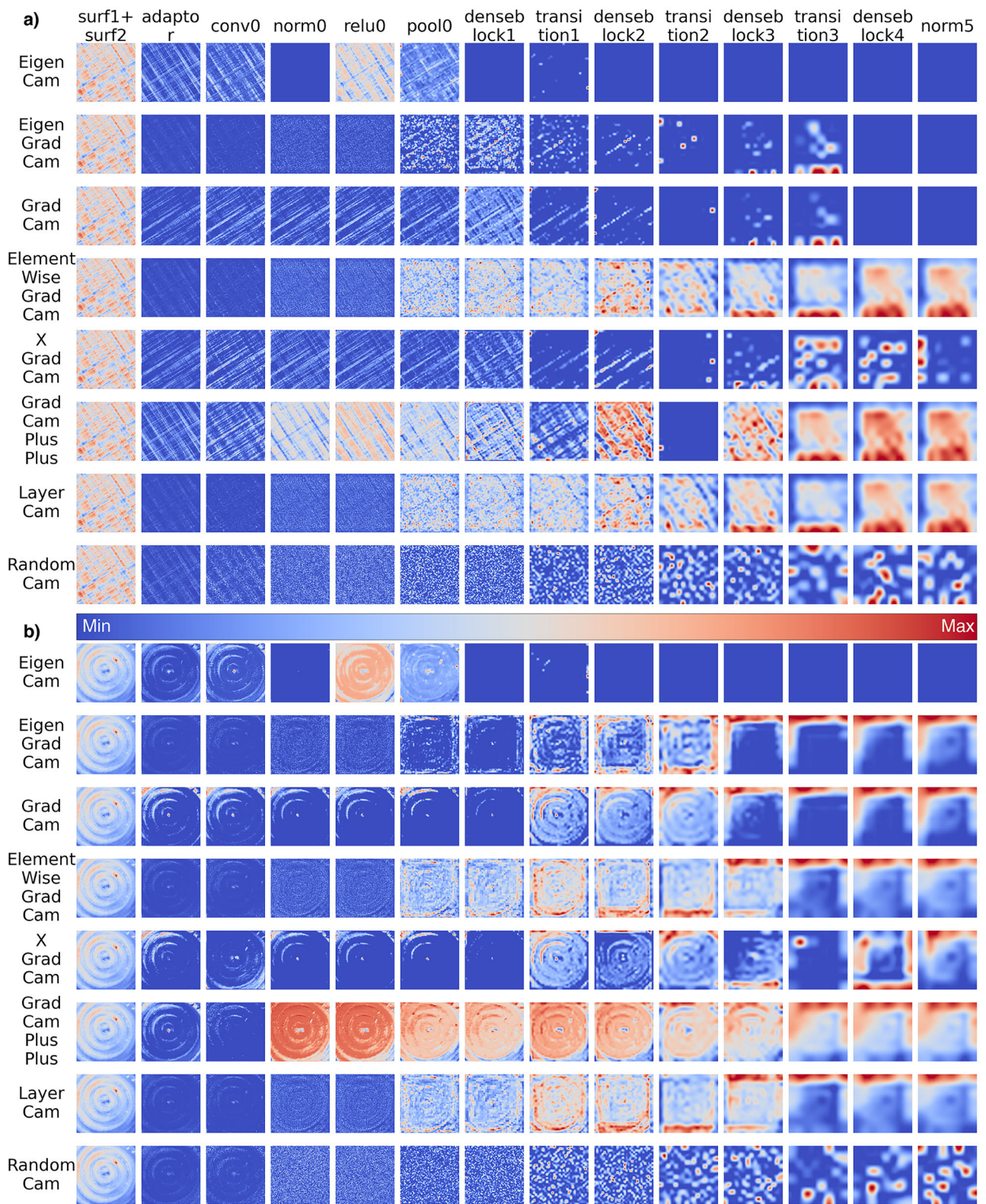
This study has several important limitations. First, the fractal theory-generated surface patterns in the training dataset are overly homogeneous, which may cause the neural network to overfit these specific surfaces. This can lead to suboptimal feature extraction, with the model either overemphasizing some surface traits or neglecting others. Second, the feature set is limited to pressure parameters and surface topography metrics, restricting the model's predictive ability to a single material type. Extending the framework to other materials or conditions would require new datasets and retraining. Additionally, the model processes pressure inputs and surface features separately, which may reduce its ability to capture interactions between these factors. Experimentally, validation was limited to TCR measurements, without direct verification of predicted contact areas. Thus, contact area predictions should be viewed as approximate rather than exact. Finally, the study does not explain the methodology behind the CAM visualizations, limiting insight into the model's decision-making process.

Future research can address several key areas: dataset acquisition, feature selection, deep learning architecture development, and model visualization. For dataset acquisition, more diverse methods should be explored to enrich training data. Including gas conduction and radiation effects can improve data accuracy and comprehensiveness. Although resource-intensive, large-scale experimental measurements would provide the most robust datasets. Optimizing feature selection is also crucial for improving model versatility. Expanding features to cover additional



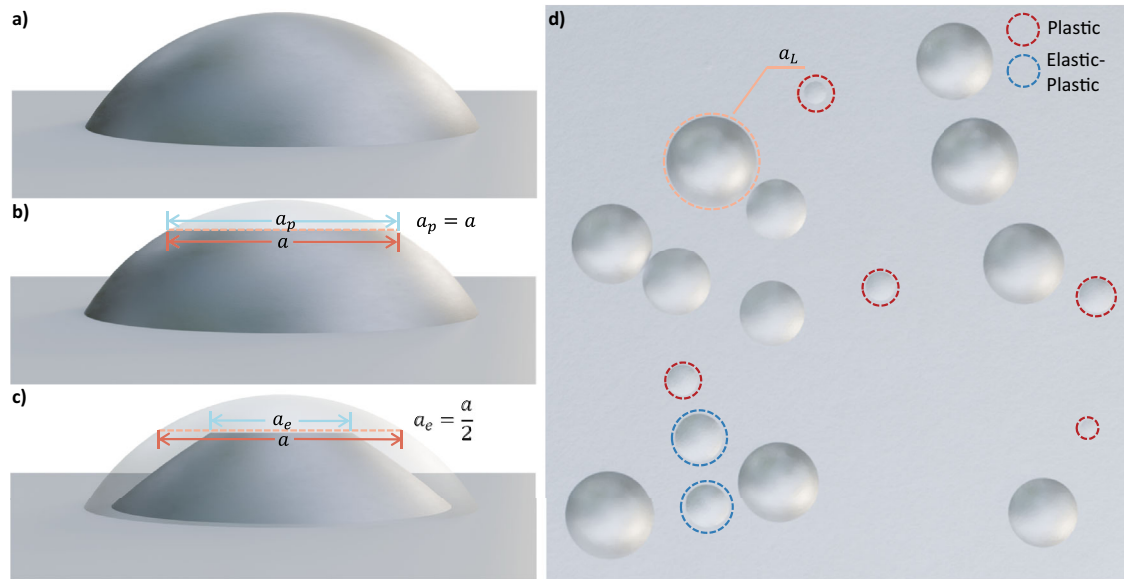
**Fig. 3 | Guided Backpropagation activation maps of the tested surfaces. a and g** show the activation maps of the original ground and turned surfaces, respectively. **b and h** depict the activation maps of the ground and turned surfaces rotated by 90°. The predicted actual contact area varies with the rotation angle at a contact pressure (P) of 1 MPa for the ground (c) and turned (i) surfaces. Similarly, the predicted thermal contact resistance (TCR) changes with the rotation angle at 1 MPa for the ground (d) and turned (j) surfaces. The predicted actual contact area fluctuates with

the sum of gradients for the ground (e) and turned (k) surfaces at 1 MPa. Finally, the predicted TCR varies with the sum of gradients for the ground (f) and turned (l) surfaces at 1 MPa. Here, Surf1 and Surf2 denote height profiles of the two contact surfaces, while Surf1 + Surf2 represents the direct summation of the height tensors of these surfaces. Similarly, Grad1 and Grad2, as well as Grad1 + Grad2, correspond to the dimensions of the Guided Backpropagation gradient matrices associated with the respective input surfaces.



**Fig. 4 | The class activation maps (CAMs) for ground and turned surfaces.** **a** CAMs for ground surfaces. **b** CAMs for turned surfaces. Here, “surf1 + surf2” represents the direct summation of the height tensors of two input surface. The names of specific layers in a deep learning model are listed at the top, while the CAM

methods are listed on the left. Each block, corresponding to a specific layer and CAM method, represents the activation extracted from that layer using the respective method.



**Fig. 5 | Schematic diagrams of the multi-spot model. a** A single spot. **b** Deformation of a single spot undergoing plastic deformation. **c** Deformation of a single spot undergoing elastic deformation. **d** Schematic illustration of deformation

types of asperities at various positions within the multi-asperity model. Elastic deformation occurs in areas not circled in the figure.

material properties and environmental variables will help the model capture broader interactions and enhance generalizability. Advancing deep learning architectures beyond CNNs is another priority. Hybrid models incorporating Transformers, diffusion models, or even large language models for semantic feature encoding could be considered, though further theoretical support is needed. Additionally, feature fusion mechanisms should be redesigned to integrate material, environmental, and surface characteristics before feature extraction, better reflecting real-world conditions where external factors influence surface behavior. Finally, improving model interpretability and visualization is essential. Employing advanced visualization techniques combined with physical theories can clarify the model’s decision-making process and build trust in its reliability.

This study aims to offer an alternative approach to research on TCR and to assist researchers in understanding the impact of surface morphology characteristics from a novel perspective.

## Methods

### Fractal surface modeling

For a specified three-dimensional fractal dimension  $D_3$  and fractal roughness  $G$ , a rough surface can be generated using the Weierstrass-Mandelbrot function<sup>27</sup>:

$$z(x, y) = L \left(\frac{G}{L}\right)^{D_3-2} \left(\frac{\ln y}{M}\right)^{\frac{1}{2}} \sum_{m=1}^M \sum_{n=n_{\min}}^{n_{\max}} \gamma^{(D_3-3)n} \times \left\{ \cos \phi_{m,n} - \cos \left[ \frac{2\pi y^n (x^2+y^2)^{1/2}}{L} \cos \left( \arctan \left( \frac{y}{x} \right) - \frac{\pi m}{M} \right) \right] + \phi_{m,n} \right\}, \quad (1)$$

where  $z(x, y)$  represents the height of the rough surface, while  $x$  and  $y$  denote the lateral position coordinates.  $L$  is the measured length of the sample, and  $M$  indicates the number of superimposed ridges. The parameter  $\gamma$  denotes the frequency density of the surface height, commonly taken as 1.5. Additionally,  $n$  represents the asperity frequency index, corresponds to the asperity size, and  $\phi$  is the random phase ranging from 0 to  $2\pi$ . The spatial frequency index  $n$  has an upper limit defined by the truncation length  $L_s$ , approximately equal to the interatomic bond length on the order of angstroms ( $\text{\AA}$ ).

### Theoretical calculation

For a single contact spot, the tip is approximated as a sphere with an apparent contact area  $a$  (Fig. 5a). Upon deformation, the contact area changes, as shown in Fig. 5b, c. As the apparent contact area increases, plastic, elastoplastic, and elastic deformation occur sequentially at the contact interface. The critical areas governing these transitions are as follows<sup>28</sup>:

$$\begin{cases} a_{c1} = \frac{G^2}{(\kappa H/2E_s)^{2(D-1)}} \\ a_{c2} = \frac{G^2}{(3H/2E_s)^{2(D-1)}} \end{cases}, \quad (2)$$

here,  $H$  denotes the microhardness of the softer material between the two surfaces,  $E_s$  represents the equivalent elastic modulus of the two materials,  $\kappa$  is the contact pressure coefficient<sup>29</sup>, and  $D$  is the two-dimensional fractal dimension<sup>30</sup>. These quantities can be calculated using the following formulas:

$$\frac{1}{E_s} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2}, \quad (3)$$

$$\kappa = 0.454 + 0.41\nu, \quad (4)$$

$$D = D_3 - 1, \quad (5)$$

where  $\nu$  denotes Poisson’s ratio.

When the apparent contact area  $a$  exceeds the critical value  $a_{c1}$ , deformation is fully elastic. when  $a$  is less than the critical value  $a_{c2}$ , deformation is fully plastic. For value of  $a$  between these two critical thresholds, deformation is elastoplastic. For a single asperity, fully plastic deformation results in a real contact area  $A_p = a$ ; whereas fully elastic deformation yields  $A_e = a/2$ . In the elastoplastic regime, the real contact area is interpolated between these two limits as  $A_{ep} = a [1 + f(a)]$ , where

the interpolation function is defined as follows:

$$f(a) = \frac{48}{441} \left[ \frac{\pi}{2} \left( \frac{a}{a_{c1}} \right)^{(2-D)/2} - 1 \right]^2 - \frac{32}{441} \left[ \frac{\pi}{2} \left( \frac{a}{a_{c1}} \right)^{(2-D)/2} - 1 \right]^3. \quad (6)$$

Correspondingly, the forces ( $f$ ) generated by a single contact spot under different contact states are<sup>31,32</sup>:

$$f_p = HA_p, \quad (7)$$

$$f_{ep} = A_{ep} \left\{ [1 - f(a)] \kappa H \frac{2}{\pi} \left( \frac{a}{a_{c1}} \right)^{(2-D)/4} + Hf(a) \right\}, \quad (8)$$

$$f_e = \frac{4}{3} \pi E_s G^{2(D-1)} a^{(1-D)} A_e. \quad (9)$$

Here  $p$ ,  $ep$ , and  $e$  denote the fully plastic, elastoplastic, and fully elastic deformation stages, respectively. For a pair of contacting surfaces with numerous asperities, the number of asperities,  $N$ , having apparent areas between  $a$  and  $a + da$  is given by ref. 33:

$$N(a) = \frac{D}{2} \frac{a_L^{D/2}}{a^{(D/2+1)}} da. \quad (10)$$

Among all asperities, the asperity with the largest apparent area is denoted as  $a_L$ . When  $a_L < a_{c2}$ , all asperities undergo fully plastic deformation. When  $a_{c2} \leq a_L < a_{c1}$ , elastoplastic deformation occurs. For  $a_L > a_{c1}$ , all three kinds of deformation are involved. These processes are illustrated in Fig. 5d. Based on the above descriptions, the real contact area  $A_r$  and the resultant force  $F$  at the contact interface under different conditions can be calculated as follows<sup>34</sup>:

$$A_r = \begin{cases} \int_0^{a_L} A_p N(a) da, & a_L < a_{c2} \\ \int_0^{a_{c2}} A_p N(a) da + \int_{a_{c2}}^{a_L} A_{ep} N(a) da, & a_{c2} \leq a_L < a_{c1} \\ \int_0^{a_{c2}} A_p N(a) da + \int_{a_{c2}}^{a_{c1}} A_{ep} N(a) da + \int_{a_{c1}}^{a_L} A_e N(a) da, & a_L \geq a_{c1} \end{cases}, \quad (11)$$

$$F = \begin{cases} \int_0^{a_L} f_p N(a) da, & a_L < a_{c2} \\ \int_0^{a_{c2}} f_p N(a) da + \int_{a_{c2}}^{a_L} f_{ep} N(a) da, & a_{c2} \leq a_L < a_{c1} \\ \int_0^{a_{c2}} f_p N(a) da + \int_{a_{c2}}^{a_{c1}} f_{ep} N(a) da + \int_{a_{c1}}^{a_L} f_e N(a) da, & a_L \geq a_{c1} \end{cases}. \quad (12)$$

The thermal contact conductance at each contact spot can be derived from heat flow tube theory<sup>35</sup>:

$$h_{stage} = \frac{2k_s a_{stage}^{1/2}}{\sqrt{\pi} [1 - (A_r^*)^{1/2}]^{3/2}}, \text{ stage} = p, ep, e, \quad (13)$$

Here, ‘stage’ refers to the mechanical contact state of the asperity.  $A_r^*$  represents the ratio of the real contact area, and  $k_s$  is the equivalent thermal conductivity. These quantities are calculated as follows:

$$A_r^* = A_r / L^2, \quad (14)$$

$$\frac{2}{k_s} = \frac{1}{k_1} + \frac{1}{k_2}, \quad (15)$$

where  $k_1$  and  $k_2$  denote the thermal conductivities of the two materials in the contact pair. By again considering the range of  $a_r$ , the thermal contact conductance for the three scenarios can be determined as

follows<sup>34</sup>:

$$H_c = \begin{cases} \int_0^{a_L} h_p n(a) da, & a_L < a_{c2} \\ \int_0^{a_{c2}} h_p n(a) da + \int_{a_{c2}}^{a_L} h_{ep} n(a) da, & a_{c2} \leq a_L < a_{c1} \\ \int_0^{a_{c2}} h_p n(a) da + \int_{a_{c2}}^{a_{c1}} h_{ep} n(a) da + \int_{a_{c1}}^{a_L} h_e n(a) da, & a_L \geq a_{c1} \end{cases}. \quad (16)$$

The measurement is assumed to be conducted under vacuum conditions, thereby eliminating the need to consider heat conduction through air gaps. Furthermore, since the temperature range is near room temperature, radiation effects are negligible. Consequently, thermal conduction constitutes the sole source of TCR in this context. Thus, the TCR is:

$$R_c = 1 / H_c. \quad (17)$$

Based on the aforementioned theory, given model inputs such as material properties, surface fractal parameters, and contact pressure, the following procedure is performed. First, a maximum apparent area  $a_L$  is initialized. Using this  $a_L$ , the force  $F$  generated by the contact between the two surfaces is calculated. This calculated force is then compared with the externally applied force  $F_{ext} = PL^2$ . If the calculated force exceeds the external force,  $a_L$  is decreased; otherwise,  $a_L$  is increased, and  $F$  is recalculated. The adjustment of  $a_L$  is achieved by multiplying it by a coefficient  $\lambda$  to accelerate convergence. When the relative difference between these two forces falls within a specified tolerance  $\epsilon = 1 \times 10^{-4}$ , the current maximum apparent area is considered the desired value. Subsequently, using this  $a_L$ , the real contact area and TCR can be calculated. The specific calculation workflow is illustrated in the Fig. 6a.

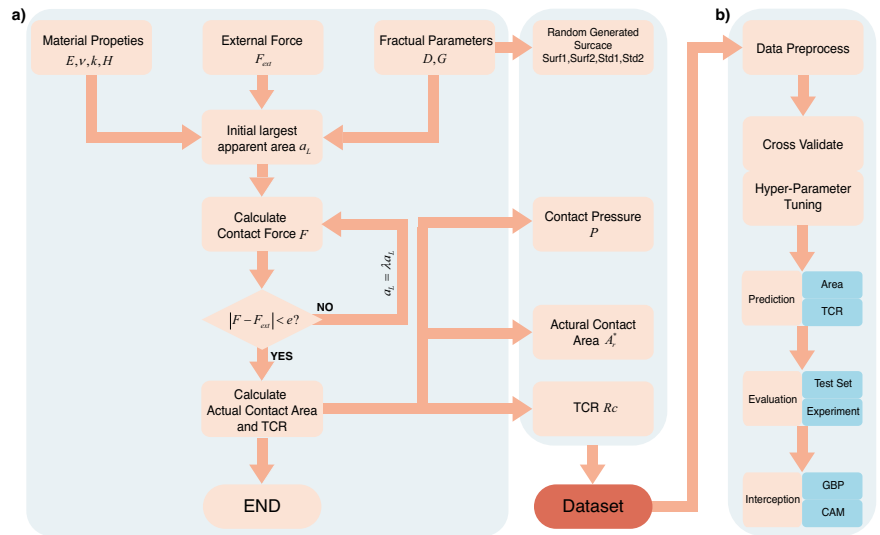
### Datasets

The data presented in this paper are derived from the theoretical model described above. Initially, fractal parameters  $D$  and  $G$  were determined, and two random surfaces characterized by  $D_1, G_1, D_2, G_2$  (calculated from  $D$  and  $G$ ) with dimensions of  $1024 \times 1024$  were generated from the fractal surface model. For theoretical models based on fractal theory both in this study and in prior research, a common practice is to approximate two rough surfaces as one rough surface and one absolute smooth surface. The generation method for the equivalent surface is detailed in the Supplementary Method 3. These fractal parameters serve as inputs to the theoretical model. Using material properties including elastic modulus, Poisson’s ratio, microhardness, thermal conductivity, and the contact pressure between the two surfaces, the actual contact area ( $A_r^*$ ) and TCR ( $R_c$ ) were calculated. The generated random surfaces, along with the input and output parameters, constitute the dataset for this study. The training set consists of 10,000 cases, while the testing set includes 2000 cases. To validate the model, experimental tests were conducted on two sets of 316 stainless steel specimens with an Ra of approximately 0.8  $\mu\text{m}$ . The mechanical properties of 316 stainless steel are summarized in Table 1. Material properties of 316 stainless steel at room temperature used in this work. TCR measurements were performed using the steady-state method<sup>36</sup>, as detailed in Supplementary Method 1 and Supplementary Fig. 4. Specific test principles and uncertainty analyses are provided in the Supplementary Method 2. The surface topography was analyzed using a surface profilometer, as described in Supplementary Fig. 3. These experimental data were also utilized for model validation.

### Selection of description set

Numerous factors influence TCR; however, not all are addressed in this study. This research primarily focuses on surface topography, rendering the topographic data of the two surfaces the most critical descriptors. Since CNNs require input data normalization, Z-score normalization is applied to the surface height data, transforming it to have a mean of 0 and a variance of

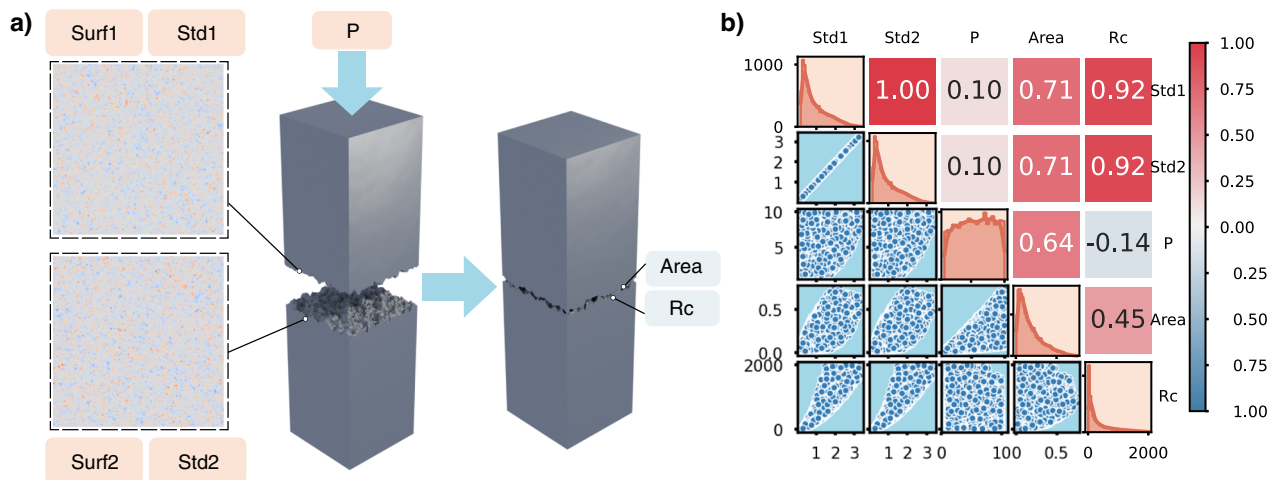
**Fig. 6 | Schematics of workflows in this paper. a.** Flowchart illustrating the theoretical model calculation and dataset generation. **b.** Flowchart depicting the processes of model training, testing, and interception.



**Table 1 | Material properties of 316 stainless steel at room temperature used in this work**

Elastic Modulus (GPa)	Poisson's ratio	Micro-hardness (MPa)	Thermal Conductivity (Wm <sup>-1</sup> K <sup>-1</sup> )
215 <sup>a</sup>	0.26 <sup>a</sup>	2200 <sup>42</sup>	15 <sup>a</sup>

<sup>a</sup>These values come from the specimen provider.



**Fig. 7 | Details of the descriptor set. a** The physical significance of each descriptor. **b** The relationships among different features. The diagonal elements display the distribution of each feature, the upper triangular section presents the correlation

coefficients along with a heatmap of the features. In contrast, the lower triangular section depicts the joint distributions of paired features. The units of the parameters are as listed in Table 2.

1 as follow:

$$z_i = \frac{z_i - \mu}{\sigma}, \mu = \sum_{i=0}^n \frac{z_i}{n}, \sigma = \sum_{i=0}^n \sqrt{\frac{(z_i - \mu)^2}{n}}, \quad (18)$$

where  $n$  represents the total number of data points in the surface height profile,  $z_i$  denotes the height at each point. Thus,  $\mu$  and  $\sigma$  correspond the mean and standard deviation of height, respectively. Notably,  $\sigma$  denotes the root mean square roughness of the surface. To preserve the integrity of surface topography information, the standard deviations of the two surfaces are included as input parameters. Consequently, a comprehensive set of descriptors for surface information includes Surf1, Surf2, Std1, and Std2. To

simplify the model, all surfaces are assumed to be composed of the same material and tested under identical environmental conditions. Therefore, material properties, ambient temperature, and ambient pressure are excluded. Only contact pressure is considered as a parameter affecting TCR. According to theory, the actual contact area between two surfaces ultimately governs the TCR<sup>6</sup>. Furthermore, the actual contact area plays a critical role in various fields, including friction, lubrication<sup>37</sup>, and coating<sup>38</sup>, establishing it as a more fundamental physical quantity than TCR. Consequently, the actual contact area should be regarded as an additional variable of interest. Because experimental measurement of the actual contact area is challenging, relying solely on it as a target variable complicates model validation. Therefore, both the actual contact area and TCR were designated as target variables. Accurate predictions of TCR are expected to yield

**Table 2 | The description set of thermal contact resistance prediction in our deep learning model**

Descriptor	Unit	Meanings
Surf1 <sup>a</sup>	-	The surface topologic of the first contact surface.
Surf2 <sup>a</sup>	-	The surface topologic of the second contact surface.
Std1	μm	The standard deviation of height of the first surface.
Std2	μm	The standard deviation of height of the second surface.
P	MPa	The contact pressure.
Area	%	The percentage of actual contact area of two surfaces.
Rc	mm <sup>2</sup> WK <sup>-1</sup>	The TCR.

<sup>a</sup>The data here has been normalized.

correspondingly reliable estimates of the actual contact area, providing additional validation. In summary, the complete set of descriptors in this study includes Surf1, Surf2, Std1, Std2, P, Area, and Rc, as is shown in Fig. 7a with their specific definitions detailed in Table 2.

The data distribution and correlation coefficients for each descriptor in the dataset. The variance of surface height ranges from 0 to 3 μm, indicating that the surface roughness is of a comparable magnitude. Pressure values span from 1 to 10 MPa, the actual contact area varies between 0 and 0.8%, and the TCR remains below 2000 mm<sup>2</sup>WK<sup>-1</sup>. The actual contact area shows strong positive correlation coefficients with Std1, Std2, and P. In contrast, the TCR exhibits a stronger positive correlation with Std1 and Std2 but no discernible correlation with pressure, which contradicts theoretical predictions. Analysis of the distribution of P relative to Std1 and Std2, as well as the distribution of P alone, reveals that P is uniformly distributed across its range. This uniformity leads to a similarly uniform distribution of TCR with respect to P, thereby obscuring any observable correlation.

**Deep learning algorithm**

Our neural network model consist of three components: an adapter, a pre-trained CNN, and a regression network. The input surface is sequentially processed by the adapter and the pre-trained CNN. Subsequently, the regression network receives the output from the pre-trained CNN along with additional parameters to perform regression and estimate the target parameters, as depicted in Fig. 8a. The pre-trained CNN employs architectures such as ResNet<sup>39</sup>, DenseNet<sup>40</sup>, and VGG<sup>41</sup>, which have demonstrated superior performance in image processing tasks. In this study, we utilized the pre-trained DenseNet121<sup>40</sup> architecture without any special consideration to accelerate model training. Given that the image input have three channels but only two contact surfaces are available, input channels of the first convolutional layer are adjusted from three to two. Furthermore, these CNN architectures are used exclusively for surface feature extraction. Since these architectures typically include classification modules, we replace the classification module with an identity layer, which outputs the inputs unchanged, ensuring that only the extracted features are forwarded to subsequent modules. The structure of DenseNet121 is shown in Fig. 8c–f.

For image input, the standard required size is 224 × 224; however, the measured surface data size is 1024 × 1024. To address this discrepancy, we introduce an adapter network preceding the pre-trained network, as illustrated in Fig. 8b. The adapter uses simple convolutional operations to resize the input from 1024 × 1024 to 256 × 256. We avoid resizing to 224 × 224 to prevent employing excessively large convolutional kernels, which could cause substantial loss of surface information. For the pre-trained network, ensuring the input size is a multiple of 32 is sufficient to preserve as much input information as possible. Although directly using 1024 × 1024 surfaces as input is feasible, it would result in an exponential increase in training time. Consequently, after thorough consideration, we adopted the current architecture.

The regression network processes the extracted surface feature tensor and input parameters. The surface feature tensor typically contains over a thousand elements, while the input parameters consist of only three elements. Directly concatenating these tensors would reduce the relative influence of the input parameters on the outputs. To address this, we first apply a linear layer to halve the length of the surface feature tensor and another linear layer to expand input parameters to a length comparable to that of the feature tensor. This standardization ensures consistency in their dimensions. Outputs of these linear layers are then concatenated and passed through two additional linear layers to regress the final target parameters: the actual contact area and TCR.

The model was trained on the training set, and its performance was evaluated on a separate test set. During training, the Adam optimizer was employed with a learning rate of 0.001, and the mean squared error loss was utilized. A 5-fold cross-validation strategy was implemented: the training set was divided into five equal subsets, with four subsets used for training and one remaining subset for validation in each iteration. This process was repeated five times, ensuring that each subset served as the validation set once. The average validation loss across all five iterations was used to assess model performance. The overall process is shown in Fig. 6b.

**Guided back-propagation**

Figure 8g schematically illustrates the data flow during neural network training. Backpropagation, the standard algorithm for training neural networks, computes the gradient of the loss function with respect to the network’s weights by propagating errors backward from the output layer to the input layer. GBP, an extension of standard backpropagation, was introduced to address limitations in generating interpretable visualizations. Its core principle is to modify the backpropagation process by retaining only positive gradients, thereby emphasizing features most relevant to the activation of a specific neuron or layer. During the forward pass, input data is processed through the network, and activations are computed for each layer. In the backward pass, GBP alters the gradient flow by allowing only positive activations to pass through ReLU layers (or similar non-linearities) during the forward pass and permitting only positive gradients to propagate backward through these layers. Negative gradients are set to zero, as is shown in Fig. 8h:

$$G = \text{ReLU}(G). \tag{19}$$

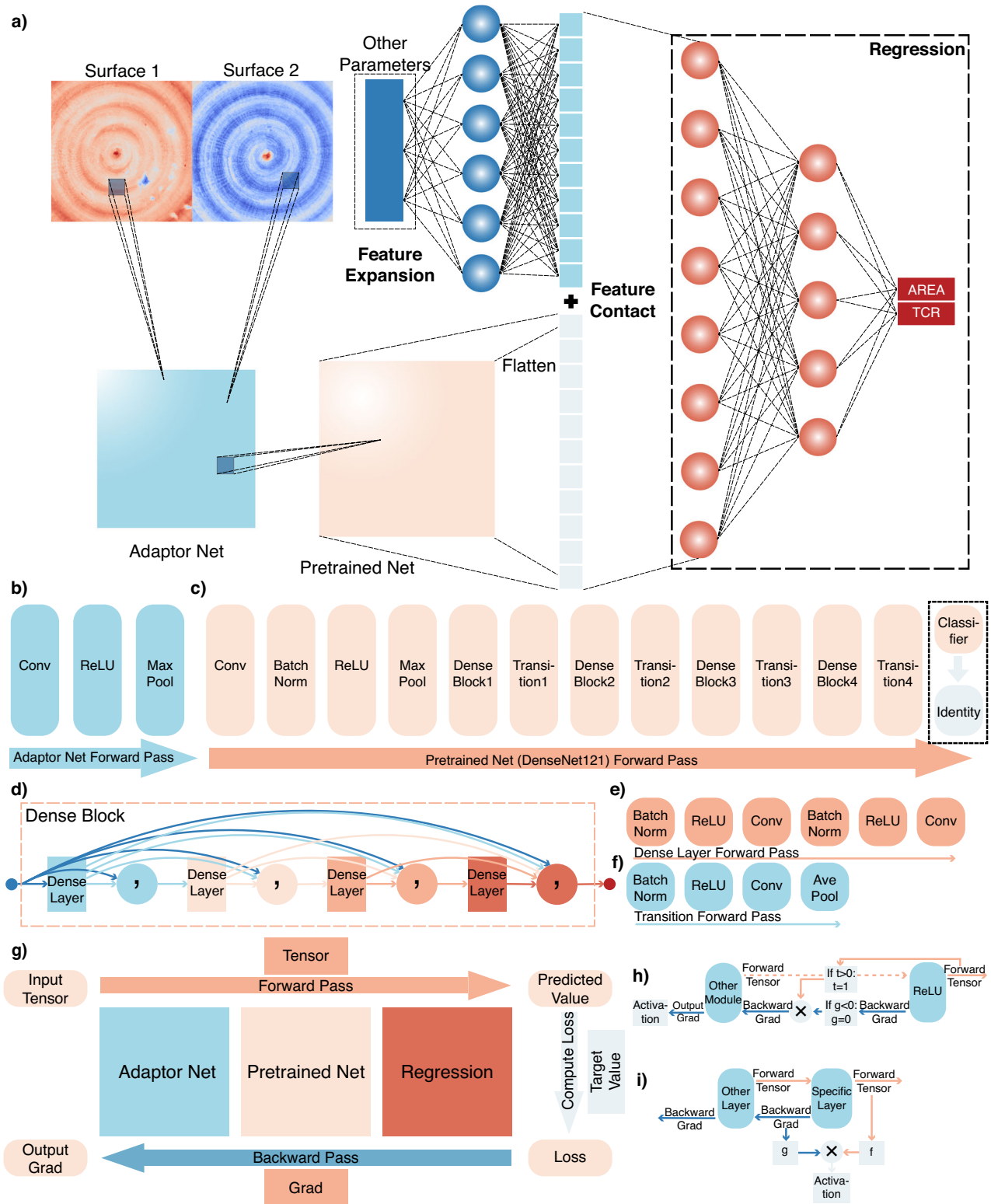
Input regions contributing greatly to the result are highlighted by the computed gradient.

**Class activation mapping**

CAMs are a fundamental technique in deep learning, employed to highlight regions of an input image that are most influential in a CNN’s classification decision. By leveraging the weights from the fully connected layer immediately preceding the output, CAM generates a class-specific activation map for classification tasks. In this work, we extend CAM’s applicability to regression problems. The activation map is mathematically defined as the weighted sum of feature maps from the last convolutional layer, where the weights correspond to the output-specific weights from the final layer:

$$A = W^T \cdot F. \tag{20}$$

Here, **W** denotes the weight vector for the specific result, and **F** represents the feature maps. In this context, **F** typically represents a function of the feature tensor during forward propagation, whereas **W** generally denotes a function of the gradient tensor during backward propagation. The specific functional form is determined by particular CAM algorithm employed, as is shown in Fig. 8i. Details of different CAMs are provided in Supplementary Method 5. By visually highlighting surface regions that contribute most substantially to predictions, this approach elucidates the model’s decision-making process, thereby enhancing CNN interpretability in real-world applications.



**Fig. 8 | Schematics of the architecture of our surface regression deep learning prediction model. a** Overall architecture of the model. **b** Basic architecture of the Adaptor Net. **c** Basic architecture of the modified DenseNet121. **d** Structure of the

Dense Block. **e** Structure of the Dense Layer. **f** Structure of the Transition block. **g** Schematic illustration of data flow in the neural network. **h** The principle of guided back-propagation. **i** The principle of Class Activation Mapping.

**Data availability**

The data that support the findings of this study are available on request from the corresponding author, [P.Z.].

**Code availability**

Codes that support the findings of this study are available in <https://github.com/Joe-zhouman/TcrSurfModel>.

Received: 27 February 2025; Accepted: 3 September 2025;  
Published online: 10 October 2025

## References

1. Mahajan, R. et al. Co-packaged photonics for high performance computing: status, challenges and opportunities. *J. Lightw. Technol.* **40**, 379–392 (2022).
2. Chen, W.-Y., Shi, X.-L., Zou, J. & Chen, Z.-G. Thermoelectric coolers for on-chip thermal management: materials, design, and optimization. *Mater. Sci. Eng. R: Rep.* **151**, 100700 (2022).
3. Lin, J., Liu, X., Li, S., Zhang, C. & Yang, S. A review on recent progress, challenges and perspective of battery thermal management system. *Int. J. Heat. Mass Transf.* **167**, 120834 (2021).
4. Zhang, Y. et al. Simultaneous electrical and thermal rectification in a monolayer lateral heterojunction. *Science* **378**, 169–175 (2022).
5. Chen, J., Xu, X., Zhou, J. & Li, B. Interfacial thermal resistance: past, present, and future. *Rev. Mod. Phys.* **94**, 025002 (2022).
6. Sun, D. et al. A review of thermal contact conductance research of conforming contact surfaces. *Int. Commun. Heat. Mass Transf.* **159**, 108065 (2024).
7. Wang, H., Liu, Y., Cai, Y., Liu, Z. & Yang, Y. Fractal analysis of the thermal contact conductance for mechanical interface. *Int. J. Heat. Mass Transf.* **169**, 120942 (2021).
8. Vu, A. T., Gulati, S., Vogel, P.-A., Grunwald, T. & Bergs, T. Machine learning-based predictive modeling of contact heat transfer. *Int. J. Heat. Mass Transf.* **174**, 121300 (2021).
9. Siddappa, P. G. & Tariq, A. Experimental estimation of thermal contact conductance across pressed copper–copper contacts at cryogenic-temperatures. *Appl. Therm. Eng.* **219**, 119412 (2023).
10. Liu, Y., Ji, Y., Ye, F., Zhang, W. & Zhou, S. Effects of contact pressure and interface temperature on thermal contact resistance between 2Cr12NiMoWV/BH137 and  $\gamma$ -TiAl/2Cr12NiMoWV interfaces. *Therm. Sci.* **24**, 313–324 (2020).
11. Pan, X. S. et al. Research progress of thermal contact resistance. *J. Low. Temp. Phys.* **201**, 213–253 (2020).
12. Cui, T. F., Li, Q. & Xuan, Y. M. Characterization and application of engineered regular rough surfaces in thermal contact resistance. *Appl. Therm. Eng.* **71**, 400–409 (2014).
13. Ma, C., Zhao, L., Shi, H., Mei, X. S. & Yang, J. A geometrical-mechanical-thermal predictive model for thermal contact conductance in vacuum environment. *Proc. Inst. Mech. Eng. Part B-J. Eng. Manuf.* **230**, 1451–1464 (2016).
14. Sun, X. G., Meng, C. X. & Duan, T. T. Fractal model of thermal contact conductance of two spherical joint surfaces considering friction coefficient. *Ind. Lubrication Tribol.* **74**, 93–101 (2022).
15. Cheng, Y., Wan, Z., Feng, X. & Long, Y. Fractal model of thermal contact conductance considering thermal stress and asperity interactions. *Int. J. Heat. Mass Transf.* **230**, 125787 (2024).
16. Dai, Y. J. et al. A test-validated prediction model of thermal contact resistance for Ti-6Al-4V alloy. *Appl. Energy* **228**, 1601–1617 (2018).
17. Dai, Y. J., Ren, X. J., Wang, Y. G., Xiao, Q. & Tao, W. Q. Effect of thermal expansion on thermal contact resistance prediction based on the dual-iterative thermal-mechanical coupling method. *Int. J. Heat. Mass Transf.* **173**, 121243 (2021).
18. Wang, C., Lin, Q., Hong, J., Zhou, Y. & Pan, Z. Effect of load application method on thermal contact resistance and uniformity of temperature distribution. *Appl. Therm. Eng.* **229**, 120625 (2023).
19. Dong, Y., Zhang, P., Chen, M. & Lian, W. Numerical study of thermal contact resistance considering spots and gap conduction effects. *Tribology Int.* **193**, 109304 (2024).
20. Zhou, T., Zhao, Y. J. & Rao, Z. H. Fundamental and estimation of thermal contact resistance between polymer matrix composites: a review. *Int. J. Heat. Mass Transf.* **189**, 122701 (2022).
21. Ren, X.-J., Gou, J.-J., Dai, Y.-J. & Tao, W.-Q. Exploring thermal contact conductance between two contact solids by artificial neural network. *Int. Commun. Heat. Mass Transf.* **136**, 106182 (2022).
22. Feng, Z. Y., Yan, J. T. & Gao, Y. W. Prediction of contact resistance between copper blocks under cyclic load based on deep learning algorithm. *Aip Adv.* **12**, 075009 (2022).
23. Xian, Y., Zhang, P., Zhai, S., Yang, P. & Zheng, Z. Re-estimation of thermal contact resistance considering near-field thermal radiation effect. *Appl. Therm. Eng.* **157**, 113601 (2019).
24. Zhou, M., Guo, P. Y., He, Z. Y., Zhang, P. & Tao, L. The thermal contact resistance dataset and the artificial intelligence-driven prediction of thermal contact resistance in multi-material systems. *Chinaxiv e-prints*, ChinaXiv:202504.200193 (ChinaXiv, 2025).
25. Springenberg, J. T., Dosovitskiy, A., Brox, T. & Riedmiller, M. In *Proc. 3rd International Conference on Learning Representations (ICLR)*. arXiv:1412.6806 (ICLR, 2015).
26. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2921–2929 (IEEE, 2016).
27. Berman, M. A. A. D. H. A multivariate Weierstrass–Mandelbrot function. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **400**, 331–350 (1997).
28. Ren, X.-J. et al. Numerical study on the effect of thermal pad on enhancing interfacial heat transfer considering thermal contact resistance. *Case Stud. Therm. Eng.* **49**, 103226 (2023).
29. Chang, Y. et al. Effect of joint interfacial contact stiffness on structural dynamics of ultra-precision machine tool. *Int. J. Mach. Tools Manuf.* **158**, 103609 (2020).
30. Chen, M., Xie, Z., Zhang, P. & Lian, W. Prediction of high temperature thermal contact conductance considering radiation effects based on fractal theory. *Tribol. Int.* **207**, 110620 (2025).
31. Greenwood, J. A., Williamson, J. B. P. & Bowden, F. P. Contact of nominally flat surfaces. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* **295**, 300–319 (1966).
32. Johnson, K. L. One hundred years of Hertz contact. *Proc. Inst. Mech. Eng.* **196**, 363–378 (1982).
33. Liu, Y., Meng, Q., Yan, X., Zhao, S. & Han, J. Research on the solution method for thermal contact conductance between circular-arc contact surfaces based on fractal theory. *Int. J. Heat. Mass Transf.* **145**, 118740 (2019).
34. Ma, C., Liu, J. & Wang, S. Thermal contact conductance modeling of baring outer ring/bearing housing interface. *Int. J. Heat Mass Transf.* **150**, 119301 (2020).
35. Mikić, B. B. Thermal contact conductance; theoretical considerations. *Int. J. Heat. Mass Transf.* **17**, 205–214 (1974).
36. Zhang, P., Xuan, Y. & Li, Q. A high-precision instrumentation of measuring thermal contact resistance using reversible heat flux. *Exp. Therm. Fluid Sci.* **54**, 204–211 (2014).
37. Bonnevie, E. D., Baro, V. J., Wang, L. & Burris, D. L. In situ studies of cartilage microtribology: roles of speed and contact area. *Tribology Lett.* **41**, 83–95 (2011).
38. Vazirinasab, E., Jafari, R. & Momen, G. Application of superhydrophobic coatings as a corrosion barrier: a review. *Surf. Coat. Technol.* **341**, 40–56 (2018).
39. He, K., Zhang, X., Ren, S. & Sun, J. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778 (2016).
40. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4700–4708.
41. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, arXiv:1409.1556 <https://doi.org/10.48550/arXiv.1409.1556> (2014).
42. Svahn, F., Mishra, P., Edin, E., Åkerfeldt, P. & Antti, M. L. Microstructure and mechanical properties of a modified 316 austenitic stainless steel alloy manufactured by laser powder bed fusion. *J. Mater. Res. Technol.* **28**, 1452–1462 (2024).

## Acknowledgements

This work is sponsored by the Guangxi Natural Science Foundation (Grant Nos. 2019GXNSFFA245003 and 2022GXNSFBA035626), the National Natural Science Foundation of China (Grant No. 52076048), and the Guangxi Key R&D Program (grant number Gui Ke AB25069193).

## Author contributions

M.Z. conceived and directed the project. M.Z. and ZY.H. prepared the datasets, developed the training architecture, and performed training. M.Z., ZY.H., PY.G., and P.Z. performed indentation, analyzed the data, and discussed the results. M.J.C. designed the validation experiments, provided critical experimental data, and performed the comparative analysis with FEM simulations. P.Z. and L.T. provided intellectual guidance, managed project finances, and reviewed the progress periodically. M.Z. and ZY.H. co-wrote the manuscript with input from all authors. All authors reviewed and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44172-025-00508-0>.

**Correspondence** and requests for materials should be addressed to Ping Zhang or Lin Tao.

**Peer review information** *Communications Engineering* thanks Mohammad Asif and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: [Rosamund Daw]. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025