FACTA UNIVERSITATIS
Series: **Mechanical Engineering** Vol. 23, No 4, 2025, pp. 757 - 785
https://doi.org/10.22190/FUME250307026J

Original scientific paper

PREDICTION OF CONTACT DISTRIBUTION ON ROUGH SURFACES USING DEEP LEARNING ALGORITHMS

Ilkwang Jang, Yong Hoon Jang

Yonsei University, School of Mechanical Engineering, Seoul, Republic of Korea

ORCID iDs: Ilkwang Jang https://orcid.org/0000-0002-4388-979X
Yong Hoon Jang https://orcid.org/0000-0002-5436-5465

Abstract. *The contact characteristics of rough surfaces play a crucial role in determining friction, wear, thermal resistance, and electrical conductivity. This study proposes a deep learning approach to efficiently predict the contact distribution of rough surfaces based on surface image information alone, and evaluates its effectiveness against numerical methods. A U-Net architecture was employed for predicting contact areas under varying scales and load conditions, using a dataset of 100,000 fractal surfaces generated via the random midpoint displacement (RMD) method. The results indicate that the deep learning model achieved performance comparable to conventional numerical methods in predicting both contact areas and electrical contact resistance, with minimal error observed in electrical contact resistance prediction. The model approached the contact prediction as an image segmentation task, enabling faster and more efficient computations than traditional numerical approaches. High performance across metrics such as Dice coefficient, Jaccard index, Bradford Factor (BF) score, and pixel accuracy highlighted its ability to maintain prediction accuracy while significantly enhancing computational efficiency. Additionally, by leveraging two-dimensional (2D) fast Fourier transform (FFT) techniques, the model effectively captured both low- and high-frequency characteristics, accurately predicting large-scale and fine-scale features of contact areas, while reducing computation time by more than 95% compared to numerical models. These findings demonstrate that the deep learning algorithms can effectively address multiscale contact problems, offering reliable data for various engineering design applications, including friction, wear, and thermal/electrical resistance, as well as enabling real-time analysis and large-scale simulations.*

Key words: *Contact distribution, Fractal rough surfaces, Deep Learning, U-Net, Scale variations, Force variations*

Received: March 07, 2025 / Accepted July 17, 2025
Corresponding author: Yong Hoon Jang
Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
E-mail: jyh@yonsei.ac.kr

1. INTRODUCTION

The interaction between asperities governs the contact behavior of rough surfaces, which typically occurs at the micrometer or submicrometer scales, resulting in the formation of distributed micro-contact regions. These micro-contact regions contribute significantly to frictional resistance by inducing disparities between the normal and tangential forces, thereby impeding relative motion. In addition, they play a crucial role in defining the electrical and thermal contact resistances, thus inhibiting conduction in both domains. Given the critical importance of friction and electrical/thermal contact resistance in engineering disciplines such as tribology and heat transfer, the accurate prediction of micro-contact regions is of substantial significance [1-5].

However, obtaining microcontact distributions experimentally poses significant challenges because of the need for sophisticated instrumentation capable of detecting minute surface deformations [6,7]. Direct measurements of the contact pressure distributions are even more demanding, rendering numerical analysis practically the only feasible approach for acquiring such detailed contact information. Over the years, many numerical models have been developed to facilitate the analysis of contact on rough surfaces, with several models demonstrating considerable efficacy in predicting contact areas [8-10].

Nevertheless, numerical methods generally require a preprocessing step in which the surface is discretized into finite elements. The resolution of this discretization is intrinsically linked to the accuracy of the contact analysis. Achieving precise results requires a fine discretization scale, which in turn significantly increases the computational costs. Multiscale approaches have demonstrated that the surface structures vary with scale, leading to the subdivision of contact areas into smaller clusters. Despite these advances, a comprehensive theoretical foundation for the convergence between the applied load and contact area in multiscale contact problems remains underdeveloped [11]. Consequently, solving contact problems involving rough surfaces necessitates the incorporation of fine surface details at smaller scales, which drastically increases the number of elements and variables in numerical models, ultimately resulting in exponential growth in computational complexity.

As an alternative approach, deep learning techniques, particularly artificial neural networks (ANNs), have garnered significant interest in recent years. In particular, convolutional neural networks (CNNs) have driven remarkable progress in image processing and analysis by excelling in learning local features and structures from images. They have demonstrated superior performance in tasks such as image classification and object detection, while also offering significant computational speed and accuracy. Despite these advancements, the application of deep learning in the field of tribology remains nascent [12-16], and leveraging deep learning to address contact mechanics problems remains relatively unexplored.

Several studies have employed ANN-based methods in contact mechanics. However, they primarily focus on predicting scalar outputs - such as the total real contact area or other mechanical parameters - using pre-calculated scalar surface topography parameters and load conditions as inputs through standard or deep ANN architectures. For example, Kalliorinne et al. developed an ANN model that takes 14 areal roughness parameters, hardness, and load as input to predict various contact mechanical responses, including the real area of contact [17]. Similarly, Suman and Prajapati proposed an optimized deep ANN model employing parameters such as root mean square roughness (Sq), skewness (Ssk), kurtosis (Sku), pattern ratio, and applied load to predict the total real contact area [18]. In contrast to these parameter-based

regression approaches, this study employs a distinct deep learning strategy leveraging CNNs to predict the contact area distribution directly from surface height images.

This study employs deep learning techniques to predict the contact area distributions on rough surfaces. By using the surface height information as input data and conceptualizing the contact state as a classification problem for each pixel, the task can be cast as an image segmentation task. CNN-based segmentation models can then be employed to predict the contact areas efficiently and accurately. Specifically, this study proposes a model based on the U-Net architecture, which has demonstrated exceptional performance in medical image analysis [19,20], to predict the distribution of contact areas on rough surfaces. This approach is designed to overcome the inherent limitations of the conventional numerical methods by providing a computationally efficient and accurate means of predicting contact distributions.

The deep learning algorithm developed in this study offers ease of application because the structures of existing datasets related to rough surfaces and contact area distributions align well with the proposed model. However, contact area distributions are inherently dependent on the spatial resolution, necessitating the acquisition of datasets across multiple scales. This issue parallels the super-resolution problem [21], a well-known challenge in deep learning research. For simplicity, we did not incorporate cross-scale correlations.

Given the rapid advancements in deep learning technologies and their potential applications in tribology, it is imperative to determine whether deep learning can effectively learn the complex nonlinear relationships inherent in contact mechanics. Should such learning prove feasible, identifying the critical factors that facilitate this predictive capability and understanding the limitations of the approach will be paramount.

Therefore, this study investigated the feasibility of employing deep learning models to replace the preprocessing and computational phases of numerical analysis for prediction of contact phenomena. Furthermore, we propose a deep learning architecture capable of capturing the geometric and contact characteristics of interacting surfaces with the ultimate goal of developing a model that can predict contact area distributions under diverse conditions and environments. Ultimately, the objective is to address the inherent complexity and convergence issues associated with numerical analysis, while offering a faster and more accurate predictive tool for contact area distributions.

2. Deep Learning Procedures

The objective of this study is to derive the contact area distribution from images of rough surfaces. To accomplish this, information regarding the contact pressure distribution and contact area distribution on rough surfaces is essential, along with a detailed methodology for constructing the dataset required to train the deep learning model. This dataset will be used for the design and implementation of the deep learning architecture to predict the contact area distribution under given contact conditions.

2.1. Rough Surface Generation

A sufficient number of rough surface contact pressure datasets are required to train the deep learning architecture. Because it takes considerable time to measure actual rough surfaces and calculate the contact pressure to construct such a dataset, rough surface and

contact pressure datasets were constructed by implementing a surface based on fractal theory and applying a method that allows quick computation to efficiently obtain the dataset.

Actual surfaces consist of rough surfaces that exhibit multiscale properties that continuously change according to the resolution of the measurement devices as surface roughness measuring instruments advance. These rough surfaces exhibit hierarchical scales that persist until a resolution limit is reached [22-24]. In this study, the rough surface was assumed to be a self-affine fractal and generated using the random midpoint displacement (RMD) method [25]. Rough surfaces were generated as 1 × 1 mm square sections, which were sufficient to capture the representative roughness characteristics. To ensure the similarity between the fractal surface and the actual surface, the fractal dimension D was limited to 2.5, and the rough surface height (h) was constrained within a specific range of 0 to 50 μm. Statistical analysis of the generated surfaces revealed an arithmetic average roughness (Ra) of approximately 7 μm, which falls within typical engineering surface roughness values. In the RMD method, the height at the midpoint between nodes defined at the previous scale is determined by a mean of zero, and the variance is influenced by the fractal dimension. According to this definition, the surface becomes rougher as the surface roughness and variance increase. In this study, a comprehensive dataset consisting of 100,000 rough surface geometries and their corresponding contact pressure distributions was constructed to train and test the deep learning model. The generated surfaces exhibited consistent statistical properties while maintaining the specified height range. Rough surfaces were obtained according to the given scale $L/2^n$, where each mesh element had a size of $L/2^n \times L/2^n$ (L = 1 mm). The multiscale surfaces produced results at scales ranging from n=5 to n=9.

2.2. Contact Analysis by Direct Numerical Simulations

A numerical analysis was required to determine the contact area distribution on rough surfaces generated using the RMD algorithm. The contact problem addressed in this study was limited to elastic bodies with frictionless contact. The findings of this study can also serve as fundamental information for extending the problem to include friction. The contact conditions for the frictionless case are as follows [26].

$$u(x,y) = \Delta - g_0(x,y) \quad (x,y) \in A \tag{1}$$

$$u(x,y) > \Delta - g_0(x,y) \quad (x,y) \notin A \tag{2}$$

$$p(x,y) > 0 \qquad (x,y) \in A \tag{3}$$

$$p(x,y) = 0 \qquad (x,y) \notin A \tag{4}$$

where

$$u(x,y) = \frac{1}{\pi E^*} \iint_A \frac{p(\xi,\eta)d\xi d\eta}{\sqrt{(x-\xi)^2+(y-\eta)^2}}, \tag{5}$$

where $g_0(x,y)$ is the initial gap, Δ is the rigid body approach, and $p(x,y)$ is the contact pressure. Eqs. (1)-(5) indicate that at the rough surface, the gap that includes the surface displacement is zero, and the contact pressure is positive. In the separation region, the gap was greater than zero, and the contact pressure was always zero. Additionally, the term $1/E^*$ represents $(1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$, where E_1, ν_1 and E_2, ν_2 are the elastic moduli and Poisson's ratios of the two contacting bodies, respectively. The sum of the contact pressures obtained from these contact conditions was in equilibrium with the applied load F.

$$\iint_A p(x,y)dxdy = F. \tag{6}$$

Deep learning improves the accuracy through repeated training based on large datasets, making it essential to secure as many datasets as possible for effective learning. However, traditional numerical methods such as finite element analysis require a significant amount of time to derive the contact area distribution on rough surfaces, making it extremely difficult to acquire sufficient datasets. To overcome this issue, this study uses numerical analysis based on the fast Fourier transform (FFT) method and multi-level multi-summation (MLMS) technique [8,26], allowing for the rapid derivation of contact area distributions and the collection of a large number of datasets. It is important to note that the dataset used for training our deep learning model relies on the results generated by this specific FFT-based MLMS numerical solver, chosen primarily for its computational efficiency and accuracy balance, particularly for concentrated contact problems. This efficiency enabled the generation of the large-scale dataset (100,000 samples) essential for deep learning within a practical timeframe. While advantageous for speed and handling non-periodic issues better than standard FFT without extensive zero-padding [26], this method, like any numerical technique, operates under specific physical assumptions (e.g., linear elasticity) and is subject to discretization effects. Consequently, the training data inherently reflects these characteristics. We acknowledge the existence of various other numerical methods (e.g., FEM, BEM, other DFFT variants) which may possess different strengths, weaknesses, and applicability limits under various conditions, as reviewed by Müser et al. [27]. The following paragraph briefly summarizes the numerical analysis method.

For a discrete system, in which the x and y planar spaces are divided into (N_x, N_y) uniform rectangular grids along each axis, Eq. (5) yields the following displacement:

$$u_{ij} = -\sum_{k=1}^{N_x}\sum_{m=1}^{N_y} K_{i-k,j-m}p_{km}\,, \tag{7}$$

where K_{ij} is the influence coefficient reflecting Eq. (5).

The coarse and fine grids are represented by (I,J) and (i,j), respectively. The contact pressure and displacement are described as follows:

$$p_{I,J} = \hat{R}^{2s}p_{ij}, \tag{8}$$

$$u_{ij} = \hat{I}^{2s}u_{IJ} + \sum_k\sum_m C_{km}^{2s}\,p_{i-k,j-m}\,, \tag{9}$$

where $\hat{I}^{2s}$ is the interpolation operator of order $2s$, and C_{km}^{2s} represents the correction coefficients. Using this relationship, the transformation between the coarse and fine grids is performed sequentially.

Meanwhile, applying the FFT technique to the continuous functions defined in the numerical analysis and their discretization yields the following result for Eq. (9):

$$\hat{u}_{hl} = \hat{K}_{hl}\hat{p}_{hl} \tag{10}$$

where $\hat{u}_{hl}$, $\hat{K}_{hl}$, and $\hat{p}_{hl}$ are Fourier coefficients corresponding to $u_{ij}, K_{ij},$ and p_{ij}. The surface nodal deflection u_{ij} can be obtained from u_{hl} using the inverse FFT.

From the above contact analysis, the contact pressure and micro-contact area distribution were obtained. The rough surface was in contact with a flat opposing rough surface, and the load range was F=0-50 N. Fig.1 shows the rough surface, contact pressure, and contact spot distributions for scales $L/2^7$ and $L/2^8$. The rough surfaces presented in

Fig. 1 are the same as those reconstructed according to each scale. As the scale decreased, the contact area formed at larger scales is divided into smaller contact spots. In addition, the contact spots tended to form clusters. The contact pressure at these spots increased as the scale decreased. Thus far, the contact area distribution under elastic contact conditions has shown a continuous reduction as the scale decreased, demonstrating non-convergent results. To address these non-convergent results, attempts have been made to account for new physical phenomena associated with scale changes, such as variations in material properties and adhesion effects [28-31]. However, in this analysis algorithm, no new physical phenomena other than mechanical behavior changes with scale were considered. Only the contact area distribution was used in the deep learning process.

2.3. Deep Learning Model: U-Net

The contact area distribution prediction problem addressed in this study is closely related to image segmentation tasks. This specifically pertains to the classification of contact and noncontact regions in the field of computer vision. The goal is to classify each pixel as either a contact or noncontact region based on the shape of the surface. We adopted the U-Net architecture to perform this task [32]. This model demonstrated excellent performance in medical image segmentation. Owing to its structural characteristics, U-Net can effectively capture the multiscale nature of contact problems, making it highly suitable for the objectives of this study.

The U-Net architecture used in this study was designed to take a rough surface geometry as the input and output the corresponding contact area distribution, as shown in Fig. 2. Specifically, the network was configured to process a 2D input image of size $N_x \times N_y$, where N_x and N_y represent the number of grid points of the rough surface in the horizontal and vertical directions, respectively. The input image size was defined to range from $2^5 \times 2^5$ to $2^9 \times 2^9$ pixels, depending on the scale.

The U-Net structure is divided into two main parts: contracting and expanding paths. These two components align well with the characteristics of the contact problems. The contracting path progressively reduces the spatial dimensions of the input image through a series of convolution operations and max pooling. This reduction process allows the network to extract high-level features [33,34]. This process is crucial for capturing the macroscopic geometry of the rough surface and overall contact pattern. Specifically, the contracting path comprises repeated applications of the following layers.

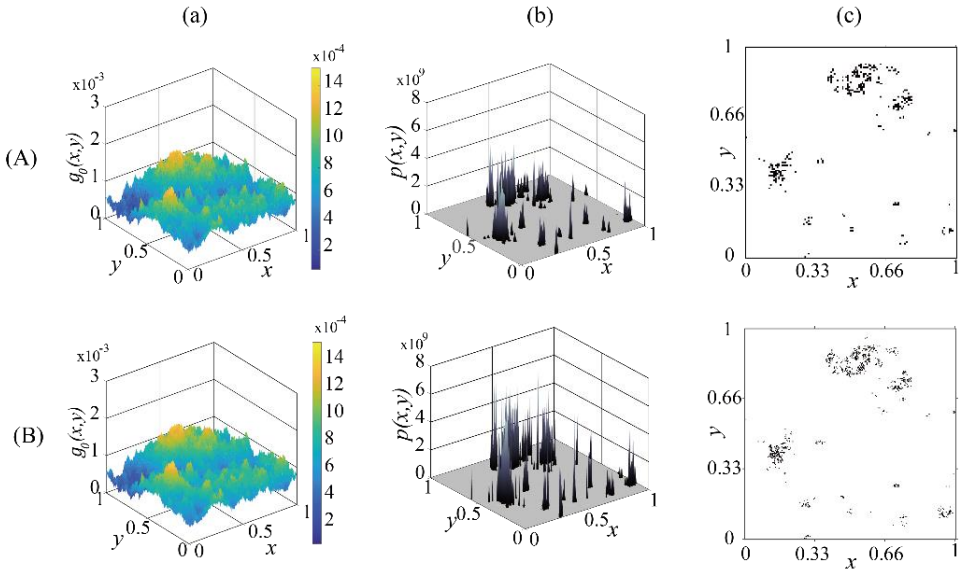

Fig. 1 Contact analysis simulation results for deep learning network training based on scale variation. Rows A) and B) show the results for scales $L/2^n$ with n=7 and n=8, respectively. For each scale, the columns represent: a) rough surface information, b) contact pressure distribution, and c) distribution of contact spots.

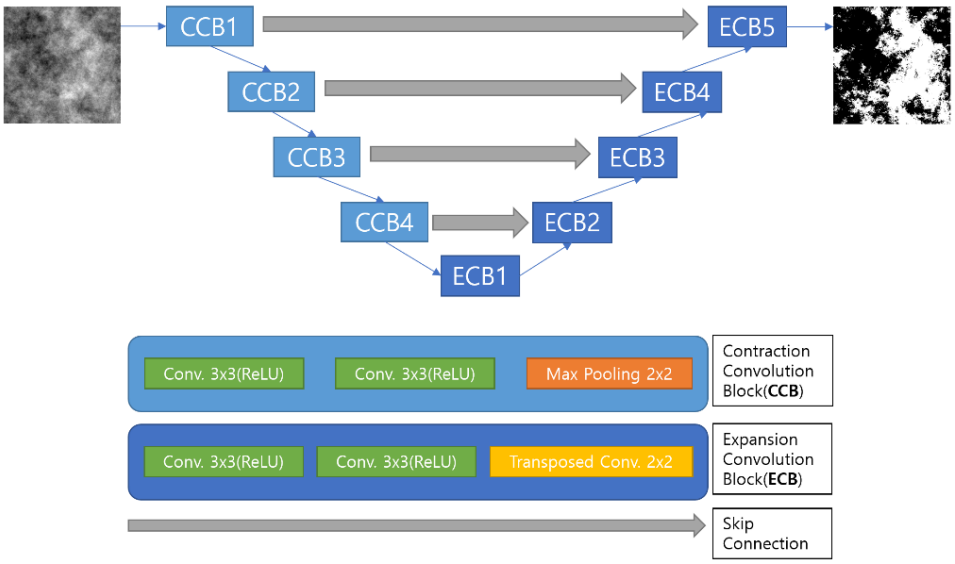

Fig. 2 U-Net architecture and network configuration for contact area prediction

The convolutional layer applies filters (or kernels) to the input data to extract features. In this study, 3 × 3 filters were used and the number of feature maps doubled after each

convolution operation. For example, 64 feature maps were generated in the first stage, 128 in the second stage, and 256 in the next stage. This process enabled the network to gradually learn more complex and abstract features.

After each convolution operation, a rectified linear unit (ReLU) activation function was applied to introduce nonlinearity. The ReLU function is defined as $f(x) = \max(0, x)$, which converts negative inputs to zero and allows positive inputs to pass. This function is widely used in deep neural networks because it is computationally simple and helps mitigate the vanishing gradient problem.

After applying the convolution and activation functions, a 2 × 2 max pooling operation was performed. This operation reduced the spatial dimensions of the feature map by half. This operation selects the largest value in each 2 × 2 region to create a new feature map. Max pooling reduces computation and increases the spatial invariance of the features. This contraction process is repeated four times to reduce the input image size to 1/16 of its original dimensions. This process allows the network to learn the macroscopic features of the rough surface and the overall contact pattern.

However, the expanding path progressively restores the spatial dimensions of the feature maps through up-sampling and convolution. It then combines the low-level features obtained from the contracting path to accurately predict the fine contact areas. The expanding path consists of the following layers.

In this study, up-sampling doubled the spatial dimensions of the feature map using the transposed convolution method. This process can be considered the reverse of max pooling, which converts low-resolution features into high-resolution features. One of the core features of U-Net is the skip connection, which is particularly important for contact problems. Skip connections help preserve important spatial information that may be lost during the down-sampling process, which is critical for accurately identifying contact areas. By linking features from the Contracting path to the expanding path, skip connections ensure that both low-level details and high-level context are available for precise contact predictions. This connection links feature maps obtained at each stage of the contracting path to the corresponding layers of the expanding path. This enables the network to leverage information about the fine surface structure (low-level features) and the overall contact pattern (high-level features) simultaneously. This is crucial for accurately identifying the boundaries of the contact areas and is essential for the precise prediction of the contact area and distribution.

A 3 × 3 convolution and ReLU activation function were applied again to the connected feature maps. During this process, the number of feature maps was gradually reduced, ultimately reaching the desired number of output channels (two in this study, representing the contact and noncontact areas). In the final layer of the Expanding Path, a 1 × 1 convolution is applied to generate the final output. This output takes the form of a probability map, in which each pixel represents the probability of belonging to a contact area.

Each convolutional layer in U-Net can be expressed as follows:

$$\boldsymbol{y}_{conv} = \sigma\big(BN(\boldsymbol{w}_{conv} * \boldsymbol{x}_{conv})\big), \tag{11}$$

where $\boldsymbol{x}_{conv}$ is the input feature map, $\boldsymbol{w}_{conv}$ is the convolution kernel, BN is batch normalization, σ is the activation function, and $\boldsymbol{y}_{conv}$ is the output feature map. The ReLU was used as the activation function and is defined as follows:

$$\sigma(\boldsymbol{x}) = \max(0, \boldsymbol{x}). \tag{12}$$

The use of the ReLU activation function allows the network to effectively model the nonlinearity of contact problems. Additionally, the application of batch normalization helps stabilize the large variability in rough surface heights and accelerates the model training process.

Each stage of the U-Net addresses specific aspects of the contact problem. The initial layers capture the fine surface structure and local features, which are essential for identifying small-scale contact points. The intermediate layers recognize mid-scale patterns and structures on the rough surface, contributing to the understanding of contact cluster formation. The deeper layers capture the overall surface geometry and large-scale contact patterns, aiding in the prediction of the overall contact area distribution. Finally, the up-sampling layers integrate these multiscale features to generate a high-resolution contact area map.

The U-Net model used in this study consists of 23 convolutional layers. This number of layers was selected to provide sufficient depth for capturing both the fine and coarse details of the rough surface, thus enhancing the ability of the model to learn the complex relationships necessary for accurate contact area predictions. This deep structure allows the network to effectively model the complex nonlinear relationship between the rough surface geometry and contact area distribution. These characteristics of U-Net make it particularly suitable for contact area distribution prediction problems. It effectively handles the multiscale nature of rough surfaces, accurately identifies the boundaries of the contact areas, and correctly predicts the overall contact pattern.

2.4. Loss Function

The choice of the loss function in the training process of a deep learning model has a significant impact on its performance and training efficiency [35]. In this study, considering the characteristics of the contact area prediction problem, a combined loss function was used that combines the binary cross-entropy (BCE) and Dice losses.

The BCE loss is one of the most widely used loss functions for binary classification problems and is defined as follows:

$$\mathrm{BCE} = -[y \log(p) + (1 - y) \log(1 - p)], \tag{13}$$

where y represents the true label (0 or 1) and p represents the predicted probability. BCE loss allows predictions to be interpreted as probabilities and is statistically well defined, because it is based on information theory and maximum likelihood estimation. It also has the advantage of providing meaningful gradients, even in extreme predictions, which stabilizes the learning process. In contact problems, the BCE loss independently calculates the loss for each pixel, enabling fine-grained predictions at the pixel level.

However, the BCE loss is sensitive to class imbalance. When the contact area occupies only a small portion of the total area, the model can predict all pixels as noncontact and achieve a low loss value. This could result in the model failing to properly predict small but important contact areas in contact problems.

To compensate for this limitation of BCE loss, Dice loss is also considered. Dice loss is based on the Dice coefficient, a metric that directly measures the similarity between the predicted and actual contact areas. To understand this metric, we introduce the concepts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These are the

basic classification outcomes defined by comparing the model predictions with the actual values. In the context of contact problems, these outcomes can be understood as follows:

1. TP: The number of pixels predicted as contact areas that are indeed contact areas. This indicates that the model correctly identified the contact areas.
2. TN: The number of pixels predicted as noncontact areas. This indicates that the model correctly identified the noncontact areas.
3. FP: The number of pixels predicted as contact areas that are actually noncontact areas. This represents a case in which the model incorrectly judged noncontact areas as contact areas, indicating an overestimation of the contact area.
4. FN: The number of pixels predicted as noncontact areas that are actually contact areas. This represents a case in which the model incorrectly judged contact areas as noncontact areas, indicating an underestimation of the contact area.

These classification outcomes are crucial for contact problems. For example, a high FP indicates that the model overestimates the contact area, predicting that contact occurs over a larger area than that in reality. Conversely, a high FN indicates that the model underestimates the contact area, predicting contact over a smaller area than that in reality. The Dice coefficient measures the overlap between the predicted and actual areas, and is calculated as follows:

$$\text{Dice Coefficient} \equiv \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{14}$$

The Dice loss is then defined as:

$$\text{Dice Loss} = 1 - \text{Dice coefficient}. \tag{15}$$

The Dice loss is less sensitive to class imbalance, making it effective in predicting small contact areas. Additionally, it is particularly useful for accurately predicting the boundaries of the contact regions, which is critical for accurately determining the contact area and distribution. Dice loss is also differentiable, making it easy to apply in the backpropagation process of deep learning models, and ranges between 0 and 1, offering an intuitive interpretation. Moreover, it automatically assigns larger weights to difficult samples (e.g., those with small contact areas), which helps address class imbalance.

To leverage the strengths of the BCE and Dice losses, this study uses a combined loss function. The combined loss is defined as follows:

$$\text{Combined Loss} = \alpha \cdot \text{BCE} + (1 - \alpha) \cdot \text{Dice Loss}, \tag{16}$$

where α is a hyperparameter that controls the balance between the two losses. The combined loss enables more balanced learning by combining the pixel-level evaluation of the BCE with the region-level evaluation of the Dice loss. The BCE captures local characteristics, whereas the Dice loss captures global characteristics, allowing optimization across multiple scales. Additionally, by adjusting the value of α, the relative importance of the two losses can be adjusted depending on the problem characteristics, which provides flexibility.

The combined use of loss-of-contact problems is particularly beneficial. Given the nature of contact problems, where both accurate boundary predictions of the contact area and an overall understanding of the contact area are essential, combining the local precision of the BCE with the global similarity measurement of the Dice loss can be highly effective.

2.5. Evaluation Metrics and Implementation

To evaluate the performance of the contact area prediction using a deep learning CNN model, the selection of appropriate evaluation metrics is crucial. In this study, the pixel accuracy, Jaccard index (intersection over union (IoU)), Dice coefficient, and Bradford Factor (BF) score were used as the main evaluation metrics [36,37]. These metrics allow the performance of the model to be assessed from various perspectives, enabling an interpretation that considers the characteristics of the contact problem.

Pixel accuracy is the most basic evaluation metric. It is calculated based on classification results, represents the proportion of correctly classified pixels out of all pixels, and is defined as follows:

$$\text{Pixel Accuracy} \equiv \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{17}$$

In contact problems, the pixel accuracy indicates the proportion of rough surfaces where contact/noncontact areas are correctly predicted, showing how well the model captures the overall contact pattern. However, this metric is vulnerable to class imbalance. For example, if the contact area occupies a very small portion of the total surface, the model can achieve high pixel accuracy even by predicting all pixels as noncontact. Therefore, it can be difficult to evaluate model performance accurately using only pixel accuracy, and this should be considered alongside other metrics.

The Jaccard index, or IoU, is calculated as the ratio of the intersection to the union of the predicted and actual areas:

$$\text{Jaccard Index} \equiv \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \tag{18}$$

This metric ranges from 0 to 1, with values closer to 1 indicating more accurate predictions. In contact problems, the Jaccard index intuitively reflects the overlap between predicted and actual contact areas, making it particularly useful for evaluating the accuracy of contact boundaries.

The Dice coefficient measures the overlap between the predicted and actual areas as defined in Eq. (14). Similar to the Jaccard index, it ranges from 0 to 1, with values closer to 1 indicating more accurate predictions. In contact problems, the Dice coefficient can be more sensitive than the Jaccard index in predicting small contact areas, making it useful for evaluating finer contact patterns.

The BF score evaluates the accuracy of the boundary region and is calculated as the harmonic mean of precision and recall:

$$\text{BF Score} \equiv \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{19}$$

where precision and recall are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{20}$$

This value represents the proportion of pixels predicted as contact areas that are actually in contact. Additionally,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{21}$$

This value represents the proportion of actual contact area pixels that the model correctly predicted as contact areas. The physical meaning of the BF score indicates the accuracy with which the model predicts the contact area boundary. This is particularly useful in problems in which the shape, such as the contact area, is important. A higher BF score indicates that the model accurately predicted the boundary of the contact area.

These evaluation metrics can be used as loss functions during the model training process or to assess the performance of the model during the validation and testing stages. Because the importance of each metric may vary depending on the nature of the problem, it is important to select appropriate metrics or consider multiple metrics comprehensively, considering the specific characteristics of the problem.

2.6. Training the Neural Network Model

In this study, a visualization process suitable for contact mechanics was conducted to utilize the fractal surfaces generated by the RMD method and the contact area data calculated by the MLMS method to train the deep learning model. The complex geometry of the rough surfaces and the resulting contact area distribution possess inherent statistical characteristics, and to effectively train the model, it is necessary to visually represent these characteristics in the data.

The process of visualizing rough surface height information and contact area distribution data was designed to preserve the precision of the numerical data while reflecting the characteristics of contact mechanics. Initially, the data were stored in a Tagged Image File Format (TIFF); however, to reduce video memory usage and image file size when representing the fine structure of the fractal surfaces, the Portable Network Graphics (PNG) format was ultimately adopted. PNG provides lossless compression and supports a 16-bit channel depth, allowing precise representation of the fine surface geometry without loss of detail.

The following steps were performed to convert the surface height information into images. First, the fractal surface data were normalized to the range [0, 1] based on the minimum and maximum heights. This normalization ensures that the statistical characteristics of height variation, which are crucial in contact mechanics research, are maintained while enabling a consistent comparison of all data points. The normalized values were then converted to the 16-bit integer range [0, 65535]. The reason for selecting 16-bit depth was to accurately capture the complex fine structure of the fractal surface and the corresponding contact mechanism. The converted data were generated as 16-bit grayscale images and stored in PNG format. The lossless compression of the PNG format was used to optimize the balance between file size and precision.

The contact area distribution data were converted into a binary mask based on the results calculated using the MLMS method. The locations where the contact pressure was greater than zero were defined as contact regions, whereas the other locations were defined as noncontact regions. This method precisely represented the discrete contact distribution observed in actual contact scenarios. The binary mask was saved in the PNG format, where a single-bit depth was sufficient to convey meaningful information.

This visualization process offered several advantages. First, using a 16-bit depth, it is possible to maintain high-resolution details of the fine surface geometry. In contact mechanics, fine surface variations significantly affect contact strength, contact area, and contact pressure distribution, making precise representation crucial. The use of the PNG format allows effective data compression, saves storage space, and enables fast data

loading in most deep learning frameworks. This is an important factor in the training process for large-scale surface datasets.

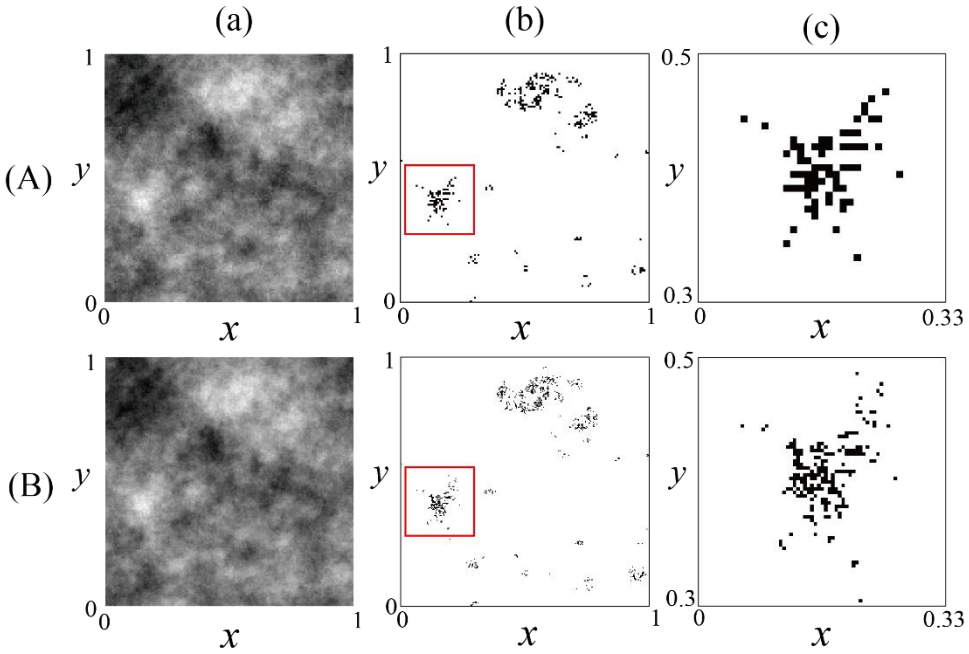

Fig. 3 Example of a preprocessed dataset for deep learning network training. Rows A) and B) show the results for scales $L/2^n$ with n=7 and n=8, respectively. In each scale, column a) shows the height information of the rough surface converted to a 16-bit grayscale PNG, column b) shows the contact area distribution represented as a binary mask, and column c) shows a zoomed-in section of the contact area distribution. This dataset includes fractal surfaces generated by the RMD method and contact area information calculated using the MLMS technique.

Moreover, visualized data are easy to analyze. Converting the surface height map and contact area distribution into images allows researchers to intuitively understand the changes in the contact conditions. This is advantageous for clearly analyzing the effects of the surface roughness or geometric changes on the contact properties. In addition, image-based data facilitate the application of data augmentation techniques for training deep learning models. Thus, the proposed visualization process can be considered an effective data preprocessing method that reflects the characteristics of contact mechanics.

The dataset generated through this preprocessing was directly used to train the deep learning model. A height map image was used as the model input, and the binary mask of the contact area was used as the label for the direct numerical simulation. Examples of the visualized rough surface height map and the binary mask of the contact area used to build the training dataset are shown in Fig. 3.

In this study, data augmentation techniques were applied to help the deep learning model learn the complex characteristics of contact mechanics more effectively. The shape

of the rough surface and distribution of the contact areas exhibited nonlinear and complex patterns, making data augmentation necessary to enable the model to capture this complexity. Data augmentation prevents the model from overfitting to specific data patterns and improves its generalization performance in predicting contact areas [38].

An important characteristic of the data used in contact mechanics is that they are independent of orientation. The roughness or shape of a rough surface is statistically isotropic; thus, the distribution of the contact areas also exhibits orientation-independent characteristics. Accordingly, flipping (vertical and horizontal) and rotation transformations were applied to the images. These transformations increase the diversity of the data while preserving the physical contact properties.

Vertical and horizontal flipping was applied with a probability of 50%, and image rotations were performed at 90, 180, and 270 degrees. These transformations generate a variety of training samples without altering the statistical characteristics of the contact area distribution, contact stiffness, or contact pressure, thereby allowing the model to learn a wider range of contact scenarios. For example, eight training samples, including the original image, were generated by applying vertical flipping, horizontal flipping, and three rotations to a rough surface image. This effectively increases the dataset size by eight times, which is particularly useful when there is a limited amount of training data.

The data augmentation technique in contact mechanics helps the deep learning model effectively learn the nonlinear behavior that occurs in various contact situations. Because the irregular distribution of contact areas and the complex surface geometry pose challenges to the generalization ability of the model, data augmentation plays a key role in mitigating these issues. In real-world scenarios, various surface conditions and orientations exist, making it essential for the model to learn how to handle such variability robustly.

Consequently, through a data augmentation technique that reflects the complex characteristics of contact mechanics, the proposed model is equipped with more reliable and practical contact area prediction capabilities.

A total of 100,000 datasets containing rough surface geometries and contact area distributions were used to train the U-Net model. Of these, 80% were used as training data and 20% were used as testing data. The batch size was set to 32 during training and the model was trained for 100 epochs.

During the training process, the Adam optimizer was used to optimize the model parameters [39]. The initial learning rate was set to 0.001 and a strategy was applied to reduce the learning rate by 20% if the validation loss did not improve for 10 consecutive epochs. Additionally, an early stopping strategy was employed to prevent overfitting and determine the optimal parameters. Specifically, training was halted if the validation loss did not improve for 20 consecutive epochs, or if the learning rate dropped below 0.0001. At the end of each epoch, the model weights were saved, and the performance metrics (Dice coefficient, Jaccard index, BF score, pixel accuracy) for both training and validation were recorded.

All experiments in this study were conducted in a Python 3.11 environment and implemented using the PyTorch library (version 2.1.2). The hardware was a workstation equipped with an AMD EPYC 7502 CPU, 256 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU.

3. RESULTS

3.1. Evaluation of the Performance of the Deep Learning Model

The results of the contact area distribution predictions based on the changes in load and scale are presented. To evaluate the performance of the model, the accuracy of the results was analyzed using various previously introduced metrics.

First, the accuracy of the predictions made by the deep learning model trained at different scales was verified. Fig. 4 shows an example comparing the direct numerical model results and predictions made by the deep learning model for the contact area distribution at different scales in the test field. In this case, the applied force was 30 N. Numerical analysis results indicated that as the scale became finer, the contact area was subdivided into smaller contact regions, which subsequently formed clusters. Additionally, it shows that new contact spots emerged at coarser scales, which were not previously present.

To quantify the accuracy of the contact area predictions made by the deep learning model in the test field, evaluation metrics, such as the Dice coefficient, Jaccard index, BF score, and pixel accuracy, are presented in Table 1. These evaluation metrics represent the average values calculated from the predictions on 4,000 test datasets, which consisted of samples at various scales under a fixed applied force of 30 N. The predicted results for each scale showed a Dice coefficient of approximately 0.9156, Jaccard index of 0.8464, BF score of 0.9521, and pixel accuracy of 0.9587. These values indicate that the deep learning prediction model demonstrates excellent overall performance in segmentation and prediction. The high Dice coefficient and Jaccard index values indicate that the model accurately identified and segmented the contact area distribution. Additionally, the high BF score suggests that the model predicts the boundaries of the contact area distribution with precision, whereas the pixel accuracy indicates that there is minimal difference between the predicted and actual contact areas, because the pixels between the distribution of the deep learning model and direct numerical simulation are classified correctly.

Second, an example comparing the numerical analysis results and predictions made by the deep learning model for the contact area distribution at different loads is shown in Fig. 5. The results demonstrated that as the load increased, the contact area expanded and a larger number of new contact spots appeared across the entire contact region under higher loads than under lower loads. The metrics used to evaluate the performance of the deep learning model in predicting the contact area distribution across different loads are presented in Table 2. These evaluation metrics represent the average values calculated from predictions on 4,000 test datasets, which consisted of samples at a fixed scale of $L/2^7$ and varying loads. The predicted results for each load exhibited a Dice coefficient of approximately 0.9289, a Jaccard index of 0.8679, a BF score of 0.9864, and a pixel accuracy of 0.9948. Compared with the results from Table 1, which analyzed scale variations, the overall results were similar, indicating the excellent performance of the prediction model. Notably, the Dice coefficient tended to be relatively high when the load was $F = 30$ N.

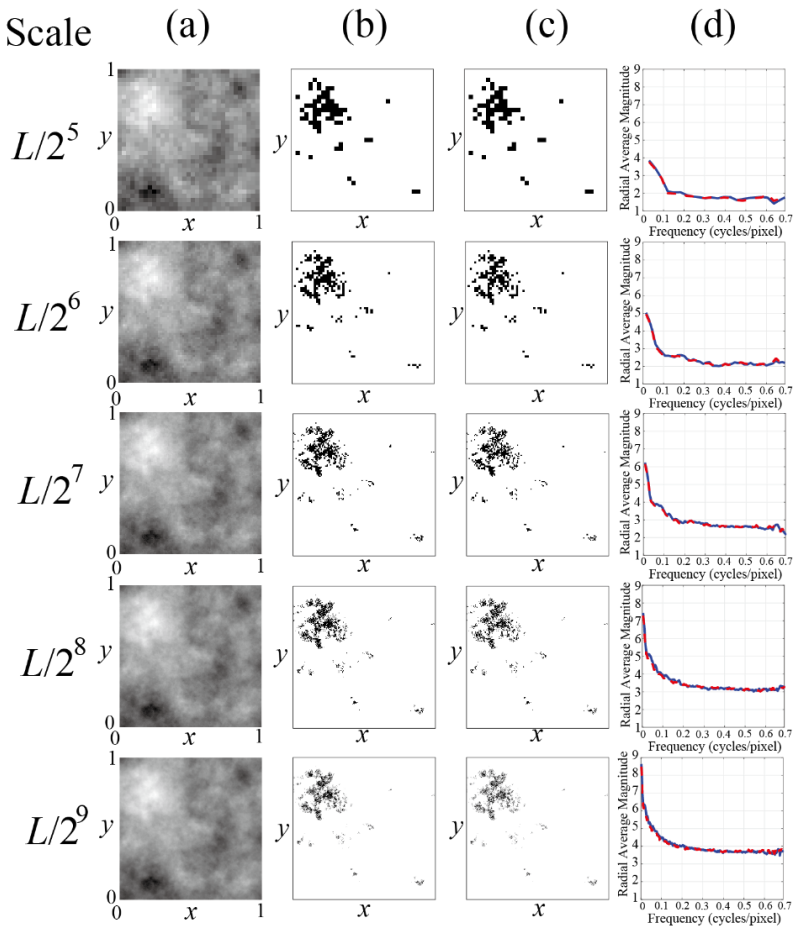

Fig. 4 Comparison of contact area distribution according to scale variation $L/2^n$, (n=5,9). Column (a) shows the rough surface image as a 16-bit grayscale PNG, (b) the contact area distribution from the direct numerical analysis, (c) the contact area distribution from the deep learning model, and (d) a comparison of radial average magnitudes by frequency (the blue line represents the numerical analysis results, and the red dashed line represents the predicted results).

Table 1 Evaluation metrics such as Dice coefficient, Jaccard index, BF score, and pixel accuracy for the predictions by the trained deep learning model at different scales

Scale	Dice Coefficient	Jaccard Index	BF Score	Pixel Accuracy
$L/2^5$	0.9169	0.8478	0.9019	0.9647
$L/2^6$	0.9129	0.8419	0.8849	0.9520
$L/2^7$	0.9188	0.8517	0.9844	0.9515
$L/2^8$	0.9138	0.8441	0.9917	0.9605
$L/2^9$	0.9154	0.8467	0.9977	0.9647

A two-dimensional fast Fourier transform (2D FFT) technique was used to analyze the structural similarity between the predicted and actual contact areas [40]. For a discrete 2D image $f(x, y)$ of size $M \times N$, the 2D FFT $F(u, v)$ is defined as follows:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}, \tag{22}$$

where $x = 0, 1, \ldots, M-1$ and $y = 0, 1, \ldots, N-1$ represent the spatial coordinates of the image, indicating the pixel positions and sampled values of the original image, respectively. In contrast, $u = 0, 1, \ldots, M-1$ and $v = 0, 1, \ldots, N-1$ are the frequency coordinates in the horizontal and vertical directions, respectively, representing the magnitude and phase of each frequency component in the transformed domain. Both sets of coordinates must be integers because this is inherent in the nature of discrete signals and the discrete Fourier transform. The periodic patterns and structural characteristics of the images were effectively analyzed by transforming the spatial domain images into the frequency domain using a 2D FFT.

The power spectrum $P(u, v)$ is computed as follows,

$$P(u,v) = |F(u,v)|^2. \tag{23}$$

To facilitate interpretation and visualization, a frequency shift operation was performed on the transformed frequency spectrum by moving the low-frequency components to the center of the spectrum. This shift can be mathematically expressed as

$$F_c(u,v) = F\left(\left(u+\frac{M}{2}\right) \bmod M, \left(v+\frac{N}{2}\right) \bmod N\right), \tag{24}$$

where the mod function is the modulus operator, which returns the remainder or signed remainder of a division after dividing one number by another.

The radial average $R(r)$ is then calculated by averaging the power spectrum over the concentric circles, as follows:

$$R(r) = \frac{1}{N_r} \sum_{u,v:\sqrt{u^2+v^2}=r} P(u,v), \tag{25}$$

where N_r is the number of frequency components at radius r, and $r = \sqrt{u^2+v^2}$ represents the radial frequency. Radial averaging reduces the two-dimensional (2D) frequency spectrum to a one-dimensional (1D) representation, allowing for a more straightforward analysis of the frequency characteristics.

Frequency analysis revealed distinct characteristics in different frequency ranges. In the low-frequency domain, the spectrum primarily reflected the overall shape and size of the contact clusters. The magnitudes of these low-frequency components are correlated with the macroscopic contact distribution. Conversely, in the high-frequency domain, the spectrum captured the fine features and boundaries of individual contact areas. As shown in Figs. 4 and 5, in the cases with high prediction performance, the frequency spectra of the predicted and actual contact areas exhibited highly similar patterns across the entire frequency range.

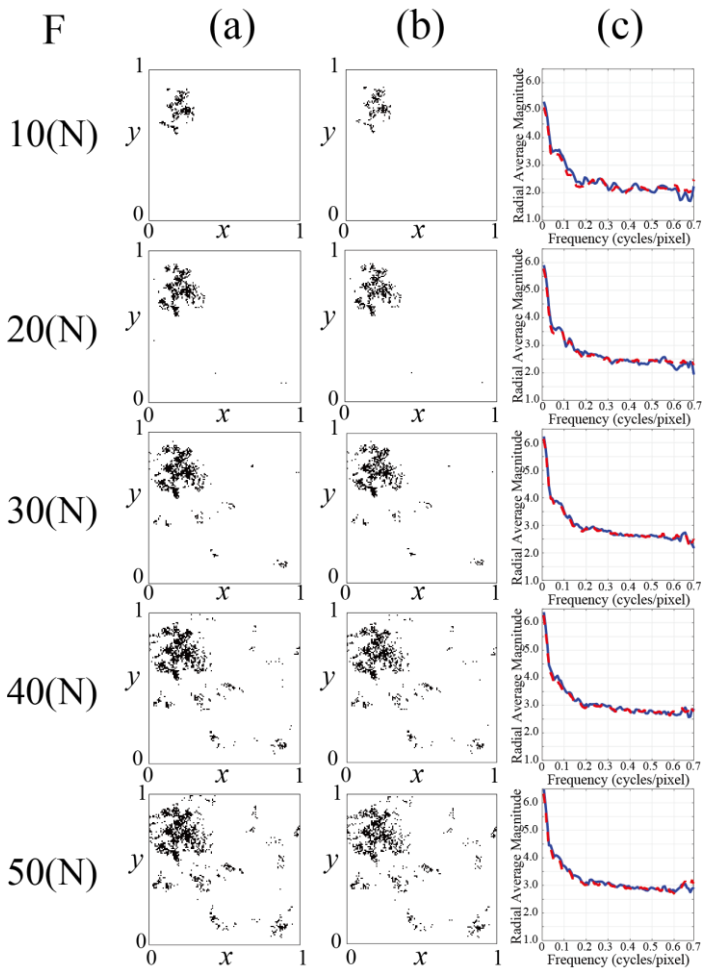

Fig. 5 Contact area distribution according to load variation. Column a) numerical analysis results, b) prediction results by the deep learning model, and c) comparison of radial average magnitudes by frequency (the blue and red dashed lines represent the numerical analysis and the predicted results, respectively).

Table 2. Evaluation metrics such as Dice coefficient, Jaccard index, BF score, and pixel accuracy for the predictions by the trained deep learning model at different load variations

Force (N)	Dice Coefficient	Jaccard Index	BF Score	Pixel Accuracy
10	0.9002	0.8186	0.9868	0.9976
20	0.9384	0.8840	0.9868	0.9971
30	0.9535	0.9112	0.9890	0.9964
40	0.9406	0.8880	0.9887	0.9938
50	0.9117	0.8377	0.9809	0.9890

The similarity in the low-frequency range indicated that the overall shape and size of the contact area were accurately predicted, whereas the match in the high-frequency range suggested that the detailed features and boundaries were precisely reproduced. This indicates that the model accurately captured both the large-scale structure and the fine details of the contact area. This similarity aligns with the Dice coefficient values in the range of 0.95-0.96 for contact area distribution across all scales, and 0.91-0.95 range for contact area distribution across all loads, as presented in Figs. 4 and 5, respectively.

To quantitatively evaluate the prediction performance of the contact area across the entire test set of 20,000 samples, we compared the actual contact areas calculated through numerical analysis with those predicted by the deep-learning model, as shown in Fig. 6. This comparison demonstrated the overall prediction accuracy and characteristics of the model. The average error rate across the complete dataset was 12.60%, with a notable trend of decreasing relative error as the contact area increased. A significant observation is that the data points are predominantly distributed below the ideal prediction line (dashed line). This indicates that the model tends to underestimate the actual contact area. The causes and implications of this systematic underprediction are discussed in detail in the following section. The dense distribution of data points around the ideal prediction line indicates consistent model performance across various contact conditions. In particular, in regions where the contact area exceeds 0.3 mm^2, the scatter between the predicted and actual values decreases, demonstrating enhanced model reliability in this range.

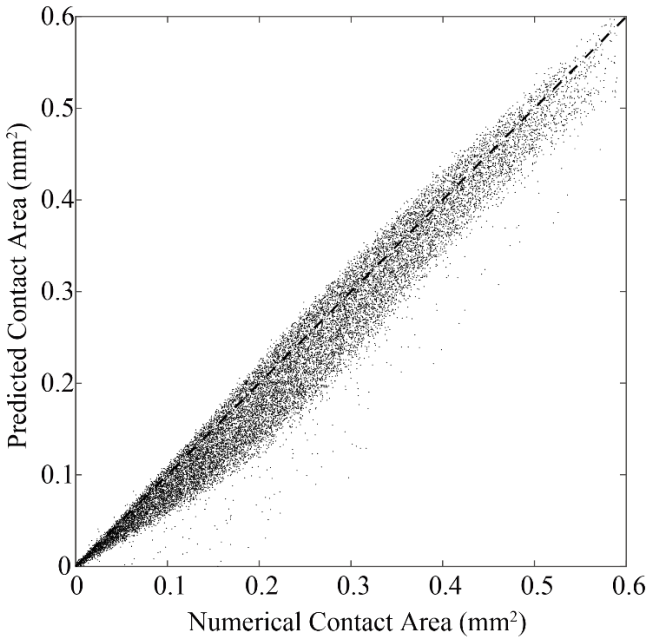

Fig. 6 Comparison of contact area between numerical analysis and deep learning predictions for the entire test dataset (20,000 samples). The dashed line represents perfect prediction.

3.2. Evaluation of the Accuracy of the Deep Learning Model from a Contact Mechanics Perspective

The elastic contact problem involving rough surfaces under normal loads typically shows that the contact area decreases, following a power law relationship, as the scale becomes finer. As shown in Fig. 7(a), the contact area decreased in accordance with the power law relationship as the scale decreased. For the rough surfaces presented in Fig. 4, the contact area predicted by the deep learning model shows a similar trend compared to the numerical analysis results, but it tends to underpredict by an average of approximately 10%.

The relationship between the applied load and contact area was also investigated. An analysis of Fig. 8(a) reveals that the deep learning model captures the fundamental characteristic of elastic contact mechanics: the linear proportionality between the normal load and the true contact area. When examining the quantitative accuracy across different load conditions (10 N-50 N), the model demonstrated consistent prediction capability while maintaining an average deviation of 11.96% from the numerical analysis results. This systematic underprediction was most pronounced at lower loads (18.14% at 10 N) and showed gradual improvement as the load increased (10.01% at 50 N), suggesting that the performance of the model is more reliable under higher load conditions where larger contact areas are formed.

This underprediction of the contact area can be primarily attributed to two factors. First, it indicates that the deep learning model misses actual contact areas more often than it mistakenly classifies noncontact areas. Second, the inherent class imbalance problem in the dataset may have led to the insufficient learning of smaller contact areas. A detailed analysis of the underprediction of the contact area is presented in the next section and in the Discussion section.

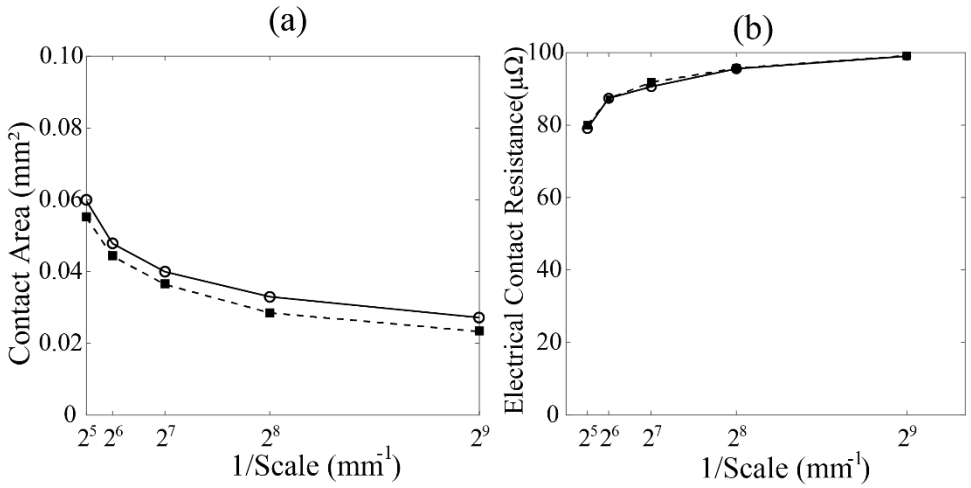

Fig. 7 Comparison between numerical analysis results and deep learning predictions of (a) variations in contact area and (b) changes in electrical contact resistance resulting from scale variations. Solid square and circle shapes represent the deep learning prediction and the numerical analysis results, respectively.

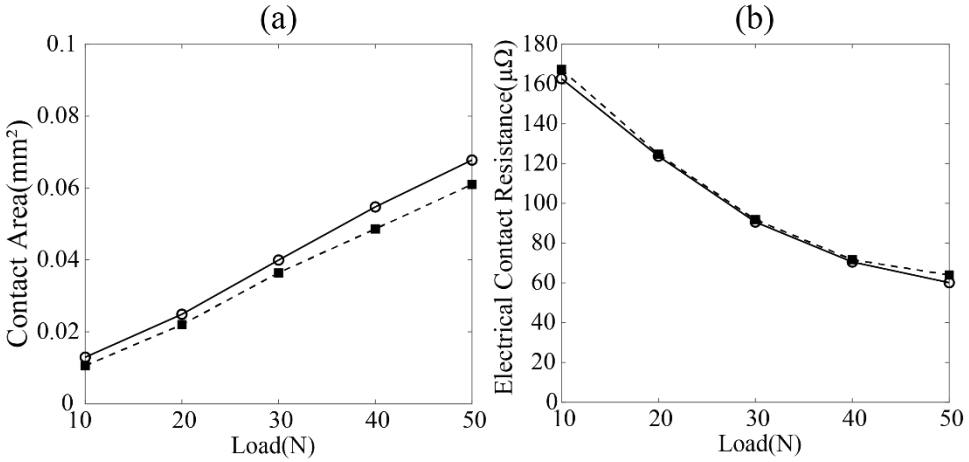

Fig. 8 Comparison between numerical analysis results and deep learning predictions of (a) variations in contact area and (b) changes in electrical contact resistance resulting from load variations. Solid square and circle shapes represent the deep learning prediction and the numerical analysis results, respectively.

Currently, the predictive performance of deep-learning algorithms is based solely on the model data generated at each scale. In other words, it does not consider causal relationships where data from a previous (coarse) scale affect data at a subsequent (fine) scale. Nonetheless, it is noteworthy that the power law relationship was maintained even in the absence of learned correlations between scale changes. This suggests that the deep learning model can implicitly learn the fundamental physical characteristics of the contact mechanism.

The electrical contact resistance (ECR) was calculated at each scale using the predicted contact area distribution. While various theoretical formulations for ECR exist, such as those developed by Paggi and Barber [41], the method presented by Boyer was selected for this study due to its suitability for our pixel-based contact analysis. Specifically, the analytical model from Section II of Boyer [42], which addresses clean metallic contact without interface films. In particular, we adopted the formulation derived for multiple square-shaped microcontacts, where the total contact resistance R is given by:

$$R \approx \frac{S(N)\rho}{nL} + \frac{\rho}{\pi n^2} \sum_{i \neq j} \sum \frac{1}{d_{ij}} \tag{26}$$

where R represents the total electrical contact resistance, and ρ denotes the electrical resistivity of the contacting material. The variable n is the number of contacting pixels, each of which is treated as a square microcontact, and L is the side length of each contact pixel. The term $S(N)$ is the shape factor for square contacts, for which we used the value 0.868, following Boyer's suggestion. The term d_{ij} denotes the center-to-center distance between the i-th and j-th contact pixels, and the summation is carried out over all distinct contact pairs.

This formulation reflects both the parallel resistance of uniformly sized microcontacts (first term) and the additional resistance arising from spatial interactions among them

(second term). All required values—such as the number of contacts n, side length L, and pairwise distances d_{ij}—were extracted directly from the predicted binary contact masks.

As shown in Fig. 7(b), the computed ECR increases and converges as the scale decreases. Despite an average deviation of approximately 10% in the predicted contact area relative to the numerical reference, the resulting ECR exhibited a discrepancy of less than 0.5%. This small difference is attributed to the fact that ECR is predominantly governed by long-wavelength surface features, which are well captured by the model, as further supported by the spectral analysis in Section 3.3.

Similarly, the electrical contact resistance was analyzed for different load conditions. Fig. 8(b) shows that the electrical contact resistance decreased monotonically with increasing load, following an inverse relationship characteristic of elastic contact mechanics. The deep learning model accurately captured this trend, maintaining a mean error of just 2.61% across all load conditions. Interestingly, the model performance showed varying accuracy levels at different loads, with the lowest error of 0.82% at 20 N and a maximum error of 6.42% at 50 N. This pattern differs notably from the contact area predictions, where higher loads generally yield better accuracy.

This phenomenon aligns with the fact that the electrical contact resistance is determined by the long wavelength roughness [43]. The current analysis of deep learning models indicates that the prediction of contact points owing to short wavelength asperities is inaccurate, leading to an underestimation of the contact area. However, the distribution of the contact surfaces influenced by long wavelength roughness was captured with a high degree of accuracy. Despite the relatively large error observed in the contact area prediction, the model consistently showed low prediction errors for the electrical contact resistance across various scales and load conditions. This suggests that the skip connection feature of the U-Net architecture effectively contributes to learning the spatial characteristics of contact regions.

Finally, the computation times of the direct numerical model and deep learning algorithm were compared. Given that the deep learning algorithm requires a very short computation time for a single surface image, the comparison was performed using 100 datasets at a scale of $L/2^9$, including the time required to store the contact distribution results. The total computation time for the direct numerical model was approximately 11,835 s, whereas the deep learning algorithm required only approximately 297 s. On average, the computation time per dataset was approximately 118 s for the direct numerical model and 3 s for the deep learning algorithm. This demonstrates the significantly higher computational efficiency of the deep learning algorithm.

3.3. Presentation of Poorly Matching Cases

The contact area distribution predicted by the deep learning model often closely matched the results of the numerical analysis; however, in some cases, the prediction led to relatively large errors in localized regions. These mispredictions tended to increase, particularly when the contact area was relatively small. Because different evaluation metrics highlight various aspects of performance, it is challenging to assess a model using a single metric. However, this explanation focuses on Dice coefficients.

Figs. 9(A)(a), 9(B)(a), and 9(C)(a) show the contact area distribution obtained through the numerical analysis, whereas Figs. 9(A)(b), 9(B)(b), and 9(C)(b) display the contact area distribution predicted using the deep learning model. By comparing and analyzing these

three sets of images, we aimed to evaluate the performance and limitations of the contact area prediction model.

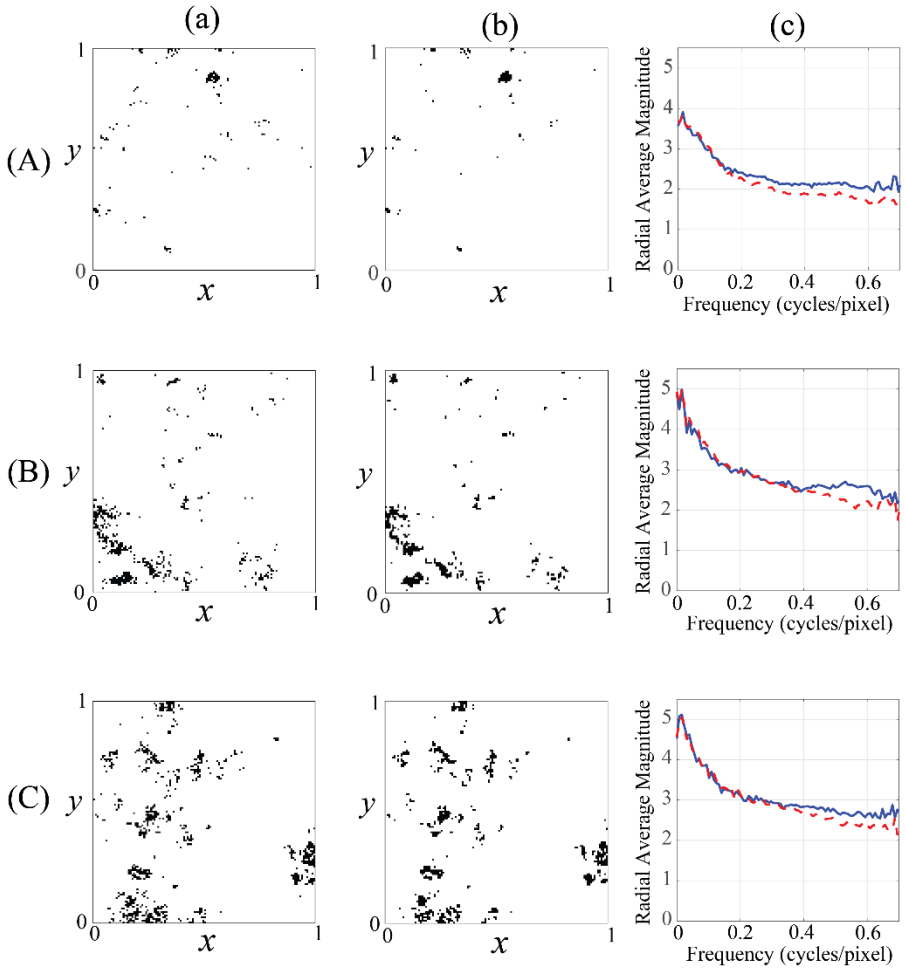

Fig. 9 Three types of contact area distributions showing poor predictions (A, B, and C). (a) results of the direct numerical model, (b) predictions by the deep learning model, and (c) a comparison of radial average magnitudes by frequency (the blue line represents the numerical analysis results, and the red dashed line represents the predicted results).

In Fig. 9(A), the numerical analysis results show a sparse distribution of contact points, whereas in the predicted image, some regions exhibit either an underestimation or overestimation of contact points. Upon analyzing the actual data for this case, the TP was 0.0054, FN was 0.0031, FP was 0.0009, and TN was 0.9907. This indicates that approximately 0.63% of the total pixels were predicted as contact areas (TP + FP), which is slightly lower than the actual contact area (TP + FN ≈ 0.85%). The fact that the FN value

is approximately 3.4 times higher than the FP value suggests that the model tends to miss the actual contact areas. This demonstrates the limitations of the model in accurately capturing fine contact points when the contact area is relatively small. The Dice coefficient for this distribution was 0.7303, indicating a relatively low prediction accuracy. In the comparison of the radial average profile, the predicted image shows a behavior similar to that of the original image in the low-frequency range, but it fails to fully capture the fine characteristics of the actual contact area in the high-frequency range. The actual contact area is 0.0085 mm^2, whereas the predicted contact area is 0.0062 mm^2, indicating a significant error of 26.62%. This substantial error quantitatively demonstrates the significantly degraded prediction performance of the model for very small contact areas. In particular, the fact that the FN value was approximately 3.4 times higher than the FP value suggests that the model struggles to capture fine contact spots.

In Fig. 9(B), the analysis results show that the TP was 0.0213, FN was 0.0087, FP was 0.0068, and TN was 0.9632. In this case, about 2.81% of the total pixels were predicted as contact areas (TP + FP), which is relatively similar to the actual contact area (TP + FN ≈ 3%). However, both FN and FP values were considerably high, indicating that the model struggled to accurately identify the contact areas. Notably, the FN value was approximately 1.28 times higher than the FP value, indicating that the model tends to miss the actual contact areas. This suggests that the model was inconsistent in predicting the varying sizes and densities of the contact points observed in the numerical analysis results. Nonetheless, the Dice coefficient for this distribution was 0.7332, indicating a slightly higher prediction accuracy than in the first case. This suggests that, as the overall size of the contact area increases, the prediction performance of the model improves to some extent. In the radial average profile, the predicted values match the original to some degree in the low-frequency range; however, as the frequency increases, the discrepancy increases, and the model fails to sufficiently capture the boundaries and fine details of the predicted contact points. In this case, the actual contact area was 0.0300 mm^2, and the predicted contact area was 0.0281 mm^2, showing a relatively small error of 6.50%. This indicates improved model prediction accuracy as the contact area increases. However, the high FN and FP values indicate that the model continues to struggle to precisely identify contact locations.

In Fig. 9(C), the analysis results show that the TP was 0.0344, FN was 0.0137, FP was 0.0078, and TN was 0.9440. In this case, about 4.22% of the total pixels were predicted as contact areas (TP + FP), which is slightly lower than the actual contact area (TP + FN ≈ 4.81%). The FN value was approximately 1.76 times higher than the FP value, indicating that the model tended to underestimate the actual contact areas. This suggests that the model fails to accurately capture the dense cluster-like contact areas observed in the numerical analyses. However, the Dice coefficient for this distribution was 0.7616, showing a relatively higher prediction accuracy compared to the previous two cases. In the radial average profile, discrepancies with the original are observed in both the low- and high-frequency ranges; however, overall, this case shows an improved performance compared with the previous two. In this case, the actual contact area was 0.0482 mm^2, whereas the predicted contact area was 0.0422 mm^2, with an error of 12.29%. Despite having the largest contact area among the three cases, the considerable error suggests that the prediction performance of the model may degrade in contact regions with complex cluster structures.

A common issue observed across all three cases was that the FN rate was higher than the FP rate, and the frequency analysis results showed a lower prediction accuracy in the high-frequency range. This suggests that the model tends to underestimate the actual

contact areas and has difficulty accurately capturing fine contact points or complex cluster structures. The frequency-domain analysis clearly demonstrates that although the model captures the overall contact pattern (low-frequency components) to some extent, it struggles to reproduce the fine structures and boundaries (high-frequency components) accurately. In particular, the decreased performance of the model in predicting low-density contact distributions or complex cluster structures raises the possibility that these extreme cases have not been sufficiently learned. The frequency analysis further highlighted these limitations. To identify the root causes of these issues, it is necessary to go beyond merely considering the model structure or quantity of training data to analyze the inherent characteristics of contact mechanics and the essential nature of the data involved.

3.4. Discussion

The deep learning model developed in this study has shown significant success in predicting the contact area distribution in contact mechanics problems. However, it also revealed several important limitations and areas for improvement. Notably, issues stemming from the characteristics of the training dataset were found to have a substantial impact on model performance.

The most prominent issue is the severe class imbalance inherent in the contact data. In most cases, the number of pixels corresponding to the noncontact areas far exceeded those in the contact areas. This "intrinsic imbalance" can cause the model to become excessively biased toward the majority class (noncontact areas), resulting in a significant reduction in the prediction accuracy for the minority class (contact areas). According to Buda et al. [44], class imbalance in deep neural networks negatively affects the model performance, and this effect becomes more severe as the imbalance ratio increases.

This class imbalance is further compounded by the characteristics of learning algorithms. The commonly used cross-entropy loss function assumes a balanced dataset and may not perform optimally under severe imbalance [45]. Moreover, in batch-based learning, if the minority class is not sufficiently represented in the batch, the model may fail to effectively learn the characteristics of that class.

To address these issues, various approaches can be considered [46]. However, the unique nature of contact mechanics problems must be taken into account. Common techniques used in classification problems, such as oversampling or undersampling, are not easily applied to contact area data owing to their specific characteristics. Simply increasing or decreasing the contact area data can distort the physical meaning. Instead, algorithm-level approaches may be more appropriate.

For example, a weighted loss function that assigns greater weight to the minority class (contact areas) can be considered. This would guide the model to pay more attention to the contact areas, which would help alleviate the class imbalance issue. To address this challenge, we employ a combined loss function. By integrating two different loss functions and adjusting the weighting parameters, the model focuses on mitigating the impact of class imbalance. Optimizing these weighting parameters plays a crucial role in improving the prediction accuracy of the model, particularly in scenarios with low contact loads. However, despite these efforts, the class imbalance issue has not been fully resolved and performance degradation is still observed in cases of extreme imbalance. Therefore, further development of more effective loss functions and additional training strategies are necessary.

In addition, strategies that consider class balance when constructing batches can be modified and applied. For example, constructing batches by sampling image patches with smaller contact areas more frequently could help strengthen the information about minor contact areas during the learning process while preserving the physical meaning of the data.

A two-stage learning approach could also be considered. In the first stage, the model would learn to extract features from the entire image. In the second stage, it would focus on predicting the contact areas based on the extracted features. This method would allow the model to learn both the overall surface geometry and the localized contact information in a balanced manner. Moreover, adding physical constraints from contact mechanics to the loss function could be another potential improvement [47]. For instance, using the relationship between the predicted contact area and actual load as a constraint could lead to physically valid predictions. This hybrid approach, which combines data-driven learning with physics-based modeling, offers a method for mitigating class imbalance while preserving physical relevance.

Future research should empirically evaluate the effectiveness of these methods and develop extended deep learning algorithms tailored to contact mechanics problems. In particular, expanding the model to predict contact area distribution (a binary classification problem) and contact pressure distribution (a regression problem with continuous values) will be essential. Overcoming the intrinsic differences between these two problems requires the development of a new architecture. One promising approach is the application of generative adversarial networks (GANs). Through competitive learning between a generator and a discriminator network, GANs can generate high-quality images [48]. By applying GANs to contact-pressure distribution prediction, it is expected that a physically valid and detailed pressure distribution can be generated. This could provide a more realistic and accurate pressure distribution than simple regression models.

Additionally, the approach developed in this study could be extended to predict surface wear. Wear is a result of the complex interaction between contact-pressure distribution and relative motion, which leads to changes in the surface geometry over time. It would be possible to use a deep learning model to predict wear patterns over time by taking the initial surface geometry and motion conditions as input. This would be a significant application in the field of tribology, contributing to the prediction of machine part lifespan and optimal design.

Furthermore, enhancing the generalization capability of the model developed in this study is another important task. Although the current model was trained under specific load conditions and surface characteristics, it should be able to operate under a wider range of conditions for practical engineering applications. To achieve this, transfer-learning techniques can be employed to develop a generalized model for various contact conditions.

Moreover, concerning the model's generalization capabilities and extensibility, a significant advantage of the proposed image-based deep learning framework lies in its potential independence from the specific numerical simulation method used for data generation. As the model fundamentally learns the mapping from a surface height image (input) to a contact area distribution mask (output), datasets generated using alternative well-established computational methods—such as FEM (e.g., using commercial software like ABAQUS or ANSYS), BEM, or other advanced simulation techniques including various FFT-based approaches and Green's Function Molecular Dynamics—could theoretically be utilized for training, provided the data is formatted into the consistent image/mask structure. Future work could explicitly investigate training and evaluating the model on datasets derived from these diverse numerical sources. Integrating data from multiple methods might even enhance the model's robustness and generalizability, potentially making it less sensitive to the inherent assumptions or artifacts of

any single simulation technique and thus increasing its practical value across varied engineering applications.

Finally, improving the interpretability of deep learning models is a critical area for future research. Current deep learning models are often regarded as "black boxes," making it difficult to understand how they make decisions. In contact mechanics, it is important to understand the decision-making process of the model, and applying Explainable Artificial Intelligence (XAI) techniques would allow for the interpretation and visualization of the decision processes of the model [49].

These research directions will expand the scope of deep learning applications in contact mechanics and contribute to the development of more accurate and reliable prediction models. This, in turn, will enable more efficient solutions to complex contact problems and significantly improve the engineering design and optimization processes.

4. CONCLUSIONS

This paper presents a deep learning algorithm for predicting the contact area distribution of rough surfaces using only surface image data. The main findings of this study are as follows:

1. The performance of the deep learning model was evaluated under various scale variations and loading conditions to predict the contact area distribution of rough surfaces. The results demonstrated that the deep learning model could accurately predict the changes in the contact area owing to variations in scale and load, and showed results similar to those of the numerical analysis for predicting electrical contact resistance. This indicates that the deep learning model effectively learned the spatial distribution of the contact points and accurately predicted the contact characteristics.

2. Using a U-Net-based deep learning model, the contact area of rough surfaces was approached as an image segmentation problem, which overcame the high computational complexity of traditional numerical methods and enabled rapid and efficient prediction. The model exhibited high performance in various evaluation metrics such as the Dice coefficient and Jaccard index. These results confirm that the deep learning model can achieve a prediction accuracy comparable to that of numerical methods while significantly enhancing computational efficiency.

3. The deep learning model demonstrated results similar to the actual data in both the low- and high-frequency domains using the 2D FFT technique, indicating that the model effectively learned the geometric characteristics of the contact phenomena. The model showed the ability to accurately predict not only the large-scale structure of the contact area but also its fine features. These findings imply that the deep learning model can adequately reflect the contact characteristics across different frequency domains.

4. The deep learning model significantly improved computational efficiency compared to conventional numerical models, reducing computation time by more than 95% for fine-scale datasets.

The deep learning algorithm developed in this study demonstrated the potential to efficiently predict complex and fine contact areas when applied to contact mechanics problems and significantly reduced computational complexity compared to conventional numerical methods. This algorithm is expected to provide reliable data for various engineering design issues, including friction, wear, and thermal/electrical contact resistance. Moreover, by reducing the computation time and cost, the proposed method has potential for application in real-time analysis and large-scale simulations across diverse engineering fields.

Acknowledgment: *The authors are pleased to acknowledge support from the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) (Grant No. 2021R1A2C3010731).*

REFERENCES

1. Greenwood, J.A., 1967, *The area of contact between rough surfaces and flats*, Journal of Lubrication Technology, 89(1), pp. 81-87.
2. Jackson, R.L., Streator, J.L., 2006, *A multi-scale model for contact between rough surfaces*, Wear, 261(11-12), pp. 1337-1347.
3. Hyun, S., Robbins, M.O., 2007, *Elastic contact between rough surfaces: Effect of roughness at large and small wavelengths*, Tribology International, 40(10-12), pp. 1413-1422.
4. Jang, Y.H., Barber, J.R., 2003, *Effect of contact statistics on electrical contact resistance*, Journal of Applied Physics, 94(11), pp. 7215-7221.
5. Bahrami, M., Culham, J.R., Yovanovich, M.M., 2004, *Modeling thermal contact resistance: A scale analysis approach*, Journal of Heat Transfer, 126(6), pp. 896-905.
6. Prevost, A., Scheibert, J., Debrégeas, G., 2013, *Probing the micromechanics of a multi-contact interface at the onset of frictional sliding*, The European Physical Journal E, 36, 17.
7. Ren, Z., Li, T., Ju, Y., Xie, H., Zhang, Q., 2024, *Characterization of the nonuniform friction behaviors in rough surfaces for the initial slip by the micro-contact and deformation zones*, Optics and Lasers in Engineering, 183, 108503.
8. Polonsky, I.A., Keer, L.M., 1999, *A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques*, Wear, 231(2), pp. 206-219.
9. Popov, V.L., Heß, M., 2015, *Method of dimensionality reduction in contact mechanics and friction,* Springer, Heidelberg.
10. Wang, Q., Sun, L., Zhang, X., Liu, S., 2020, *FFT-based methods for computational contact mechanics*, Frontiers of Mechanical Engineering, 6, 61.
11. Vakis, A.I., Yastrebov, V.A., Scheibert, J., et al., 2018, *Modeling and simulation in tribology across scales: An overview*, Tribology International, 125, pp. 169-199.
12. Ules, T., Haselmann, M., Grießer, M., Gruber, D.P., 2022, *Finger contact area analysis with convolutional neural networks*, Applied Artificial Intelligence, 36(1), 1987035.
13. Silva, A., Lenzi, V., Pyrlin, S., Carvalho, S., Cavaleiro, A., Marques, L., 2023, *Deep learning approach to the texture optimization problem for friction control in lubricated contacts*, Physical Review Applied, 19(5), 054078.
14. Feng, Z., Yan, J., Gao, Y., 2022, *Prediction of contact resistance between copper blocks under cyclic load based on deep learning algorithm*, AIP Advances, 12(7), 075009.
15. Vu, A.T., Gulati, S., Vogel, P.A., Grunwald, T., Bergs, T., 2021, *Machine learning-based predictive modeling of contact heat transfer*, International Journal of Heat and Mass Transfer, 174, 121300.
16. Lee, S., Kim, T., Lee, S., Hong, S.H., 2023, *Novel method for measuring a wear scar using deep learning*, Tribology International, 190, 109043.
17. Kalliorinne, K., Larsson, R., Pérez-Ràfols, F., Liwicki, M., Almqvist, A., 2021, *Artificial neural network architecture for prediction of contact mechanical response*, Frontiers in Mechanical Engineering, 6, 579825.
18. Suman, S., Prajapati, D.K., 2025, *Predictive modeling of real contact area on rough surfaces using deep artificial neural network*, Journal of Tribology, 147(11), 111501.
19. Shen, D., Wu, G., Suk, H.I., 2017, *Deep learning in medical image analysis*, Annual Review of Biomedical Engineering, 19, pp. 221-248.
20. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sànchez, C.I., 2017, *A survey on deep learning in medical image analysis*, Medical Image Analysis, 42, pp. 60-88.
21. Wang, Z., Chen, J., Hoi, S.C.H., 2021, *Deep learning for image super-resolution: A survey*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), pp. 3365-3387.
22. Mandelbrot, B.B., 1982, *The fractal geometry of nature*, Freeman, San Francisco.
23. Lopez, J., Hansali, G., Le Bosse, J.C., Mathia, T., 1994, *Caracterisation fractale de la rugosite tridimensionelle d'une surface*, Journal de Physique Archives, 4, pp. 2501-2519.
24. Majumdar, A., Bhushan, B., 1995*, Characterization and modeling of surface roughness and contact mechanics,* in: Bhushan, B., (Ed.), *Handbook of Micro/Nano Tribology*, CRC Press, New York, pp. 109-165.
25. Russ, J.C., 1994, *Fractal surfaces*, Plenum Press, New York.

26. Polonsky, I.A., Keer, L.M., 2000, *Fast methods for solving rough contact problems: A comparative study*, Journal of Tribology, 122(1), pp. 36-41.
27. Müser, M.H., Dapp, W.B., Bugnicourt, R., et al., 2017, *Meeting the contact-mechanics challenge*, Tribology Letters, 65(4), 118.
28. Gao, Y.F., Bower, A.F., 2006*, Elastic-plastic contact of a rough surface with Weierstrass profile*, Proceedings of the Royal Society A, 462, pp. 319-348.
29. Heß, M., Giesa, P.L., 2024, *An application of the Mossakovskii-Jäger procedure for solving plane strain adhesive contact problems of power-law graded elastic solids*, Spectrum of Mechanical Engineering and Operational Research, 1(1), pp. 64-81.
30. Popov, V.L., 2024, *Designing surface profiles with zero and finite adhesion*, Spectrum of Mechanical Engineering and Operational Research, 1(1), pp. 82-89.
31. Ciavarella, M., McMeeking, R.M., Cricrì, G., 2023, *On the afferrante-carbone theory of ultratough tape peeling*, Facta Universitatis-Series Mechanical Engineering, 21(4), pp. 727-735.
32. Ronneberger, O., Fischer, P., Brox, T., 2015, *U-Net: Convolutional networks for biomedical image segmentation*, Proc. MICCAI 2015: 18th International Conference, Munich, Germany, pp. 234-241.
33. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018, *Unet++: A nested U-Net architecture for medical image segmentation*, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support - DLMIA ML-CDS 2018, Granada, Spain, pp. 3-11.
34. Lu, H., She, Y., Tie, J., Xu, S., 2022, *Half-UNet: A simplified U-Net architecture for medical image segmentation*, Frontiers in Neuroinformatics, 16, 911679.
35. Jadon, S., 2020, *A survey of loss functions for semantic segmentation*, Proc. 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, pp. 1-7.
36. Muller, D., Soto-Rey, I., Kramer, F., 2022, *Towards a guideline for evaluation metrics in medical image segmentation*, BMC Research Notes, 15(1), 210.
37. Rainio, O., Teuho, J., Klén, R., 2024, *Evaluation metrics and statistical tests for machine learning*, Scientific Reports, 14(1), 6086.
38. Shorten, C., Khoshgoftaar, T.M., 2019, *A survey on image data augmentation for deep learning*, Journal of Big Data, 6(1), 60.
39. Sun, H., Zhou, W., Yang, J., Shao, Y., Xing, L., Zhao, Q., Zhang, L., 2024, *An improved medical image classification algorithm based on Adam optimizer*, Mathematics, 12(16), 2509.
40. Mwema, F.M., Akinlabi, E.T., Oladijo, O.P. 2019, *Two-dimensional fast fourier transform analysis of surface microstructures of thin aluminium films prepared by radio-frequency (RF) magnetron sputtering*, in: Awang, M., Emamian, S.S., Yusof, F. (Eds.), *Advances in Material Sciences and Engineering*, Springer, Singapore, pp. 239-249
41. Paggi, M., Barber, J.R., 2011, *Contact conductance of rough surfaces composed of modified RMD patches*, International Journal of Heat and Mass Transfer, 54(21), pp. 4664-4672.
42. Boyer, L., 2001, *Contact resistance calculations: Generalizations of Greenwood's formula including interface films*, IEEE Transactions on Components, Packaging and Manufacturing Technology, 24, pp. 50-58.
43. Persson, B.N.J., 2022, *On the electric contact resistance*, Tribology Letters, 70, 88.
44. Buda, M., Maki, A., Mazurowski, M.A., 2018, *A systematic study of the class imbalance problem in convolutional neural networks*, Neural Networks, 106, pp. 249-259.
45. Niaz, N.U., Shahariar, K.M.N., Patwary, M.J.A., 2022, *Class imbalance problems in machine learning: A review of methods and future challenges*, ICCA '22: Proceedings of the 2nd International Conference on Computing Advancements, Dhaka, Bangladesh, pp. 485-490.
46. Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., Japkowicz, N., 2024, *The class imbalance problem in deep learning*, Machine Learning, 113(7), pp. 4845-4901.
47. Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F., 2022, *Scientific machine learning through physics-informed neural networks: Where we are and what's next*, Journal of Scientific Computing, 92, 88.
48. Jain, A., Anusree, H., Manikandan, J., 2022, *Design and evaluation of GAN based regression model*, Proc. 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, pp. 1-6.
49. Barredo-Arrieta, A., Díaz-Rodríguez, N., Del-Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020, *Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, Information Fusion, 58, pp. 82-115.