ACTA IMEKO
ISSN: 2221-870X
September 2025, Volume 14, Number 3, 1 - 12

Deep learning and image processing-based method for automatic estimation of metal-machined surface roughness grades

Chi-Ngon Nguyen[1], Huu-Phat Tran[2]

[1] *Faculty of Automation Engineering, College of Engineering, Can Tho University, Khu II, Đ. 3 Tháng 2, Xuân Khánh, Ninh Kiều, Cần Thơ, Vietnam*
[2] *Department of Mechanical Engineering, College of Engineering, National Central University, No. 300, Zhongda Rd, Zhongli District, Taoyuan City, Taiwan 320*

ABSTRACT
Surface roughness is one of the critical technical requirements of precision machining engineering. Traditional assessment methods, such as standard sample comparisons or contact roughness measurement devices, have long shown limitations. The case of this study is in the context of machining workshops equipped with computer numerical control systems (CNC), where turning and milling methods account for an average of 60 % of the machining process. Based on convolutional neural networks and image processing techniques, this study proposes a method and a hardware structure to support non-contact roughness grade evaluation through surface texture images. The device is suitable for medium- and small-sized machine parts, meeting the practical production context of CNC machining workshops and the ISO 1302:1992 roughness grade classification standard. The training data were generated from images of surfaces with known roughness levels within the *Ra* 0.4–3.2 µm. The model achieved an average accuracy of 85.83 %, indicating the feasibility of applying convolutional neural networks and image processing to determine and assess the quality of machined surfaces.

Section: RESEARCH PAPER

Keywords: convolutional neural networks; surface roughness grade; CNC machining texture; image pre-processing; non-contact roughness measurement

Citation: Ch.-Ng. Nguyen, H.-Ph. Tran, Deep learning and image processing-based method for automatic estimation of metal-machined surface roughness grades, Acta IMEKO, vol. 14 (2025) no. 3, pp. 1-12. DOI: 10.21014/actaimeko.v14i3.2072

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received February 10, 2025; **In final form** May 27, 2025; **Published** September 2025

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding author: Chi-Ngon Nguyen, e-mail: ncngon@ctu.edu.vn

1. INTRODUCTION

In the field of precision engineering, surface roughness directly impacts the working performance and durability of components. Many studies have confirmed the influence of surface roughness on factors such as dimensional accuracy, friction, wear, fatigue strength, and heat and electrical conductivity performance [1]–[5].

The roughness profile is a critical surface attribute that refers to the shape and spatial characteristics of the surface. It is typically determined along a longitudinal line on the surface or through a cross-sectional profile, as illustrated in Figure 1 [6]. The main approaches in measuring and evaluating surface roughness are *Ra* (mean roughness, µm), *Rz* (average height between peaks and valleys, µm), and *Rq*/*RMS* (the root mean square roughness, µm) [7], [8]. They are defined by ISO 4287: 1997 [9] and standardized for tactile measurement by ISO 3274:1996 [10].

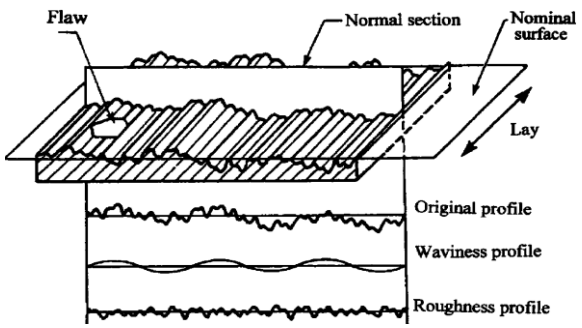

Figure 1. Surface roughness, waviness, and lay.

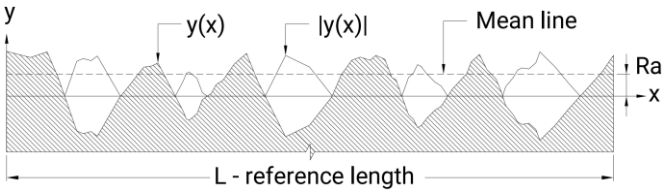

Figure 2. Ra roughness calculation method.

As the reference length increases, *Ra* tends to converge towards the true value [11], and it is also less sensitive to extreme deviations, it is computationally inexpensive, and still effectively reflects surface quality. It is also easy to visualize and interpret in measurement results.

Therefore, *Ra* is the most widely used roughness parameter in manufacturing and mass production. *Rz* and *Rq* are more sensitive to sharp surface peaks, making them more suitable for research purposes or high-resolution analysis rather than for routine use in mass production.

Currently, international standards such as ISO 4287, ASME B46.1, and JIS B0601 designate *Ra* as the primary parameter for surface roughness specification. Most surface roughness measuring instruments are configured to display *Ra* by default, further contributing to its widespread adoption.

As described in Figure 2, to simplify the computational model, this paper employs a calculation method – based on the reference length (*L*) – which removes the waviness profile using a long-wave Gaussian filter [12]. The *Ra* parameter (μm) is calculated based on the average of the absolute deviations of the surface points from its mean line over a specified reference length. The formula for calculating *Ra* is defined as follows:

$$Ra = \frac{1}{L}\int_0^L |y(x)|\,\mathrm{d}x\,, \tag{1}$$

where

- Ra is arithmetic average roughness,
- L is sampling length or cut-off length; this is the length over which the roughness value is measured and calculated,
- $y(x)$ is a function representing the surface profile at position x; it is the distance from the mean line to a point on the profile at position x,
- $d(x)$ is differential of the length x, representing a very small segment of the sampling length.

Conventional surface roughness evaluation methods, such as contact-based stylus profilometry, have long been considered the gold standard due to their high accuracy and repeatability, and comparing with a standard is also a low-cost method. However, their point-based measurement nature provides limited spatial coverage, which may not represent the surface's overall texture, while the comparison method has the limitation of low reliability, as the assessment results depend on the subjective experience of the evaluator. Surface profilometers require strict storage and operating conditions [13]; the procedure of installation and periodic calibration must be regularly performed by specialized personnel [14]. Contac t-based stylus profilometry is time-consuming, requires direct contact with the surface, and it is susceptible to wear and potential damage to both the probe and the surface being measured, particularly in precision machining contexts; if there is insufficient space to position the probe assembly properly, the probe may not make good contact with the surface in deep cavities or narrow grooves. Uncertainty due to the process and measuring instruments [15], the accuracy of this method depends on scanning speed and spacing [16], and different types of styluses can lead to discrepancies in surface roughness measurement results [17].

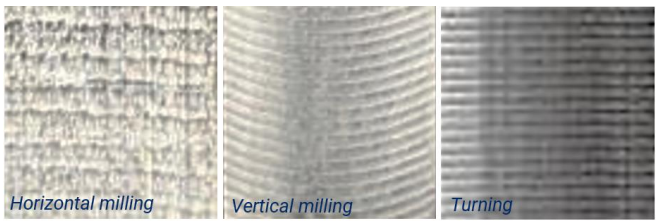

Figure 3. Surface texture is formed by different machining methods.

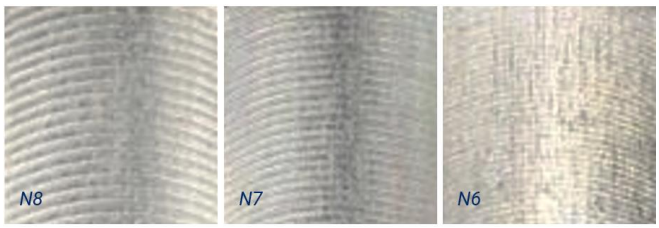

Figure 4. Different roughness grades of vertically milled surfaces.

In contrast, non-contact surface roughness evaluation using image-based deep learning offers a promising alternative. It enables rapid, non-destructive, and full-field analysis of surface characteristics without physically interacting with the material. This capability is especially critical in modern manufacturing environments that demand high throughput, inline quality control, and the handling of delicate or complex-shaped components.

As in Figure 3, we can see clear differences between surfaces in machining methods and roughness levels (Figure 4). This also forms the basis for the method of assessing roughness by comparison with standard samples. However, if the disparity is small, the method of observation and comparison with standard samples will not be feasible. A study by E. S. Gadelmawla (2010) demonstrated the correlation between *Ra* and surface texture characteristics. Accordingly, the author used texture features from the grey level co-occurrence matrix (GLCM) to estimate the surface roughness of samples machined by turning methods [18]. The relationship between GLCM texture features and surface roughness was confirmed and can be used to estimate the roughness.

Each fabrication method creates a characteristic texture on the workpiece surface. The differences again are influenced by specific cutting conditions such as the material, the cutting tools, the depth of the cut, and the cutting speed [19]–[21]. Thus, there is a close relationship between surface texture and profile, both of which directly influence and determine the surface roughness of machined components. Convolutional Neural Networks (CNN) has advantages that make it particularly suitable for recognition and classification tasks. Its structure consists of multiple layers, with convolutional layers using filters to extract features from the input data. The pooling layers reduce the spatial dimensions of the features, helping to decrease the number of parameters that need to be computed. Finally, the fully connected layers aggregate the features to perform classification or recognition [22]. This allows CNN to automatically learn and extract complex features from image data without manual intervention or specific feature programming.

In the context of the precision manufacturing industry, surface quality requirements have been emphasized [1]. Studies in recent years have clearly shown two prominent trends. The first one focuses on prediction models, in which deep learning techniques are applied to predict surface finish from input data such as process parameters, aiming at proactive surface quality control. The paper of T. Batu et al. [23] presents an overview of advances in machine learning and deep learning techniques applied by researchers. In addition, the paper discusses the limitations, challenges, and future directions for the application of deep learning networks in surface roughness prediction for additive manufacturing components. This paper clearly shows that the integration of deep learning network methods offers great potential in improving the productivity and competitiveness of additive manufacturing processes [23], [24].

Meanwhile, the other research trend focuses more on classification models, aiming to evaluate the finished surface quality. The overall goal of the studies in this group is to provide solutions for non-contact surface quality determination based on images of surface structure, and it has been yielding very clear results. This is also the research direction in which this paper will contribute additional methods and notable results in the context of CNC machining at fabrication workshops.

One of the recent studies, A. P. Rifai et al. (2020), did not extract or process structural image features. Instead, the study focused only on analysing the suitability and accuracy of five loss functions for the prediction models [25]. The predicted values obtained were compared to the actual surface roughness values measured using a stylus-based profilometer. The performance of the proposed model was evaluated for predicting the surface roughness of typical machining operations, such as outside diameter turning, slot milling, and side milling, under various cutting conditions. The models with the Mean Absolute Percentage Error (MAPE) loss function achieved the highest accuracy, as they are less affected by outlier data. The average error rate of *R*a and *R*z predictions has been confirmed to be within 10 % of the actual measured surface roughness values.

A. Giusti et al. (2020) presented a method based on CNN, where the input is a small square image representing a small portion of the surface, and it returns the *R*a value of that part of the surface [26]. The CNN is first trained using a set of numerous training samples, in which each sample is a pair consisting of one input image patch and the corresponding expected output, i.e., the true *R*a value of the surface visible in that patch. After training, the CNN is deployed in the Electrical Discharge Machine (EDM) and used to predict the *Ra* value for new surfaces that are not part of the training set. The paper presents extensive qualitative and quantitative experimental results for a range of different roughness values: $0.2 < Ra\ (\mu m) < 2.0$.

S. Jinzhao et al. (2022) developed a system based on the Region Proposal Network (RPN) combined with the Faster R-CNN neural network to simultaneously determine the surface roughness of multiple surface regions [27]. The surface regions with different roughness levels will pass through the Region of Interest (ROI) pooling layer and predictor to obtain the regions with workpieces in the image and determine the roughness grade of the workpieces. The experimental results of the model show an accuracy of up to 97.8 %. This method provides high measurement productivity. However, it only addresses the milling method with workpieces of simple shapes. In practice, the machining methods and workpieces have more complex shapes and structures.

S. Yu-peng et al. (2024), in a study on non-contact roughness measurement for surfaces machined by the grinding method, proposed the weighted window-enhanced sharpness evaluation algorithm based on the sharpness evaluation function to automatically extract the sharpest image. Then, a convolutional deep learning architecture, Convolutional Feature Extraction Network (CFEN), was trained on the dataset they built. The measurement results of the model in practice achieved an accuracy of 91.25 % [28]. The same target is surface-machined on a grinding machine.

C. Shang et al. (2024) [29] present a different approach. Their study established a machine vision detection method integrating neural networks and correlation functions. First, the neural network is used to classify and predict microscopic images of the part's surface, thereby identifying its roughness type. Then, a corresponding correlation function is determined based on the established roughness type, and surface roughness is calculated based on that correlation function. Experimental results show that the classification accuracy of the neural network reaches up to 95 %. These studies only focus on the group of surfaces machined by the flat grinding method with roughness in a narrow range of (0.3–1.6) μm.

Most recently, A. Arif et al. (2025) proposed a hybrid deep learning model combining CNN and Support Vector Machine (SVM), optimized using Bayesian algorithms (Bayesian-optimized CNN–SVM), for real-time surface roughness classification and prediction based on in-process images of machined Inconel 716 superalloy surfaces. The model was trained using hyperparameters optimized via the Bayesian Optimization Algorithm (BOA), achieving a classification accuracy of 92.37 %, which was further improved to 98.21 % with the integration of the SVM classifier. Additionally, the model demonstrated a prediction accuracy of 96.89 % for surface roughness values, along with high performance metrics: precision (98.02 %), sensitivity (98.40 %), and F1-score (98.20 %). When compared with state-of-the-art architectures, such as AlexNet, ResNet50, VGGNet, and Faster R-CNN, the proposed hybrid model exhibited superior performance while completely eliminating the need for human intervention in the image processing and feature extraction stages [30].

J. Meng et al. (2025) also proposed an automatic surface roughness recognition system based on deep learning, utilizing a dual-source laser speckle imaging device to improve the accuracy of real-time measurement during machining processes. The deep learning model Twins-SVT was enhanced by integrating a Spatial and Channel Attention (SCA) module, which effectively exploited the distinctive features from the speckle images. Experimental results on a large-scale dataset demonstrated that the SCA-Twins-SVT model achieved superior accuracy and generalization performance compared to mainstream deep learning architectures, particularly in classifying surface roughness machined by horizontal and vertical milling [31].

The studies mentioned in this section are limited to a specific machining case with a very narrow measurement range. Notably, the roughness values selected within the range are not defined according to a clear grading standard. In practical measurement, roughness is evaluated according to a threshold (grade), where a minimum roughness value is selected as the representative. Furthermore, focusing solely on a specialized context of machining method, material type, and overly narrow measurement range makes it challenging to meet the diverse evaluation requirements for manufacturing methods and materials in CNC fabrication workshops.

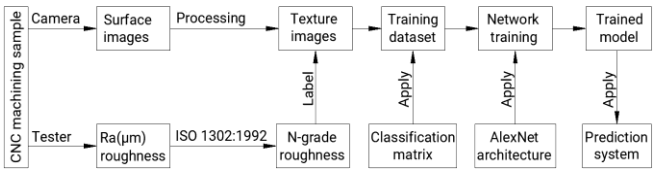

Figure 5. Proposed research framework for surface roughness classification using deep learning.

The aforementioned research gap is the focus of this study, which proposes a solution integrating image pre-processing techniques and AlexNet [32], a powerful CNN architecture for image recognition for non-contact classification of machined surface roughness grades, following the widely adopted ISO 1302:1992 standard. The flowchart in Figure 5 illustrates the main steps of the proposed methodology.

2. MATERIALS AND METHODS

2.1. Classification matrix

Within the scope of this study, the theoretical framework for the data classification method is based on two foundations: the ISO 1302:1992 roughness grade classification standard (Table 1) [33] and the roughness range that CNC machining methods (turning and milling) can achieve in typical manufacturing workshops [34]. In Table 2, the training dataset consists of 4 classes based on 4 roughness grades and 3 machining methods: turning, horizontal milling, and vertical milling.

2.2. Model selection

In this study, AlexNet was selected based on a balanced consideration of accuracy, computational cost, training time, and processing speed suitable for real-world manufacturing environments. Compared to deeper convolutional architectures, such as Visual Geometry Group (VGG) network [35] or residual network (ResNet) [36], AlexNet offers several advantages that align well with the characteristics of the dataset.

While VGG (VGG-16) contains approximately 138 million parameters, AlexNet has around 60 million. With fewer convolutional layers and larger filter sizes, AlexNet presents a more compact architecture, while retaining the ability to extract essential spatial features. This makes it particularly suitable for tasks involving medium-sized datasets and limited computational resources.

Though more advanced and featuring a much deeper network structure, ResNet demonstrates clear advantages when applied to large and diverse datasets. However, for datasets with limited size and low diversity, ResNet is more prone to overfitting. In contrast, AlexNet can achieve effective feature learning on moderately sized and diverse datasets, with a lower risk of overfitting.

2.3. Collecting and pre-processing images

AlexNet relies on image characteristics to accurately recognize and classify images. The data collection process adheres to principles ensuring representativeness for each roughness grade. This enables the CNN model to learn the surface characteristics of each machining method, as well as the distinct characteristics of different roughness levels within the same machining method. The accuracy of CNN models in classification tasks also depends on the characteristic extraction process for texture, a very important part of the training process.

Therefore, the input data needs to be protected from the effects of spatial distance changes that alter the camera's field of view, and factors causing interference, such as variations in brightness and reflection phenomena, should be minimized. The requirements mentioned above have established an important foundation for the experimental setup, especially in the design of the training data acquisition and pre-processing process and the system hardware design to ensure consistent illumination intensity throughout the image acquisition process. Figure 6 illustrates the step-by-step sequence of the image acquisition and pre-processing process, from the original image generated by the camera to the training data.

Figure 7 illustrates the hardware layout for collecting texture images. Accordingly, the distance from the camera to the surface varies depending on the size of the workpiece, resulting in different image scales. Variations in scale for texture images can significantly affect the model's training and prediction outcomes. Therefore, image scaling based on shooting distance is a crucial step for standardizing the input image ratio. This study applies an image scaling method using the zoom factor, which is calculated based on the reference distance and the measured distance obtained through a laser distance sensor. The zoom factor is determined by the proportional formula:

$$Z_{\mathrm{f}} = d_{\mathrm{n}}/d_1, \tag{2}$$

Table 1. Comparison of *Ra* (μm) and roughness grade numbers (according to table C.1 of ISO 1302:1992) [29].

N-grade	N 1	...	N 5	N 6	N 7	N 8	...	N12
***Ra* (μm)**	0.03	...	0.4	0.8	1.6	3.2	...	50

Table 2. Classification matrix by roughness grade and machining method.

N-grade	**Turning (a)**	**Horizontal milling (b)**	**Vertical milling (c)**
N5	N5(a)	N5(b)	N5(c)
N6	N6(a)	N6(b)	N6(c)
N7	N7(a)	N7(b)	N7(c)
N8	N8(a)	N8(b)	N8(c)

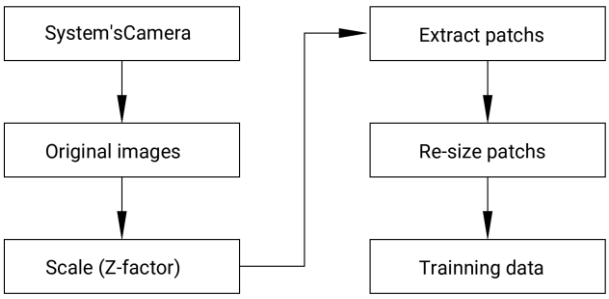

Figure 6. Flow diagram of image pre-processing and data collection for training purposes.

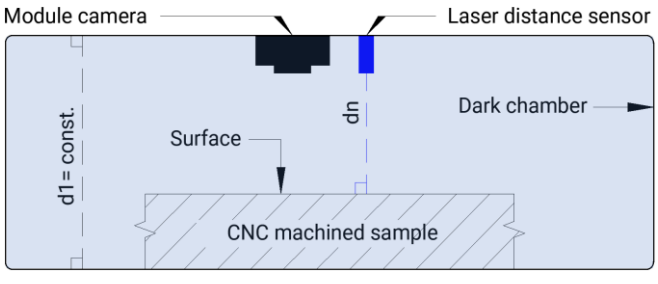

Figure 7. Hardware layout for collecting surface texture images with a laser distance sensor and camera module.

where

- Z_f is the scale factor,
- d_n is the distance (mm) obtained by the laser sensor,
- d_1 is the distance (mm) from the centre of the camera to the floor of the dark chamber and is a constant.

The size factor Z_f is applied to normalize the input image scale, ensuring that every image in both the training dataset and real-world testing is rescaled to the same ratio, which assures consistency in the ratio across images captured at different distances. Thus, an input image will be resized as follows:

$$\begin{cases} W_z = W_o \times Z_f \\ H_z = H_o \times Z_f \end{cases}, \tag{3}$$

where

- W_o, H_o are the original width and height of the original image, respectively,
- W_z, H_z are the original width and height of the new image after being multiplied by the Z_f factor, respectively.

In the process of collecting data for training purposes, the patch extraction technique helps increase the size of the training dataset. In the condition that input images are large, but the number of samples is limited, dividing a large image into multiple small patches improves the efficiency of data collection for training. A larger number of training samples enables the deep learning model to learn better and generalize more effectively. These small patches allow the model to focus on learning important local features instead of processing the entire image at once. This is particularly useful in applications requiring micro-level analysis, such as detecting scratches or cracks or evaluating surface roughness.

The AlexNet architecture applied in this study has been optimized for input images with a resolution of 227 × 227 pixels. Accordingly, a resizing filter is placed at the output of the image pre-processing stage (the input of the trained model) to adjust all images (image processing for the training dataset and image processing in the actual test) to a resolution of 227 × 227 pixels.

2.4. Training data from patch extraction

Training a convolutional neural network requires a sufficiently large number of images per class and a balance between classes to ensure the model learns the data features effectively and achieves high performance. If the number of training images is too small, the model may fail to capture the relationships in the data (underfitting), or it may overlearn the characteristics of the training data, leading to poor predictions on test datasets and real-world data (overfitting). Another key parameter to ensure compatibility with the architecture of AlexNet is that the input image size must be 227 × 227 pixels. AlexNet utilizes convolutional and pooling layers with fixed kernel sizes, so the input image size needs to be appropriate to ensure that the output after each layer is valid.

Texture images exhibit a uniform distribution of information across their entire area frame; this characteristic allows us to apply patch extraction techniques to enhance the efficiency of image acquisition and generate a sufficiently large dataset for model training. Theoretically, applying small patches with a minimum size of 227 × 227 pixels, combined with different sliding strides, allows an original image to be divided into numerous distinct sub-images, thereby increasing the number of samples for each class. The principle of this technique is illustrated in Figure 8a-b. According to block (1) of the flowchart

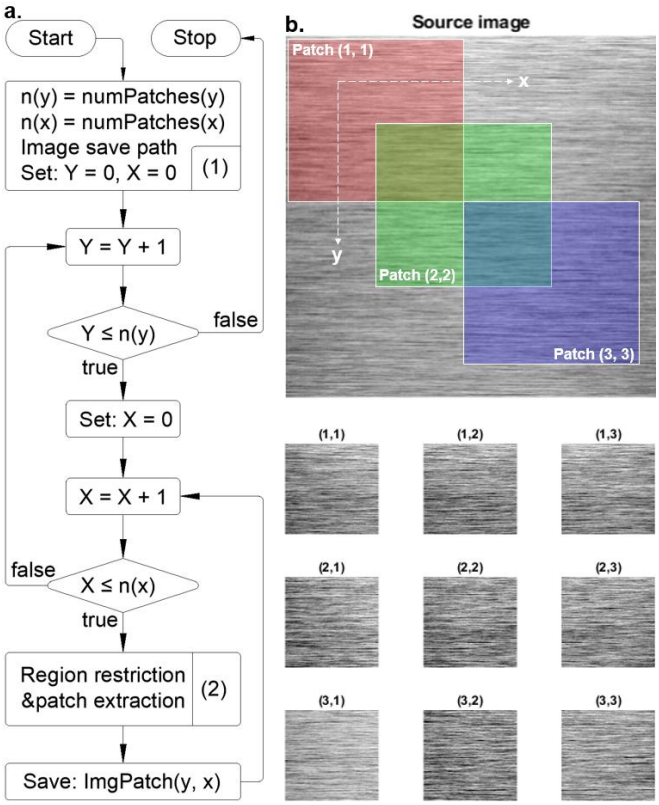

Figure 8. a) Patch extraction process, b) example showing a source texture image (537 × 537 pixels) divided into patch images (227 × 227 pixels) with a stride of 113 pixels along both x and y axes.

in Figure 8a, the total number of patches extracted from the source image can be calculated as follows:

$$Patches_{\mathrm{Total}} = n_x \times n_y \tag{4}$$

with

$$\begin{cases} n_x = \mathrm{floor}(i_x - p_x)/s_x \\ n_y = \mathrm{floor}(i_y - p_y)/s_y \end{cases} \tag{5}$$

where

- n_x, n_y are a number of patches along the horizontal and vertical dimensions of the source image,
- i_x, i_y are the width and height of the source image (pixel),
- s_x, s_y are horizontal and vertical stride lengths (pixel),
- p_x, p_y are the width and height of the patch (pixels),
- floor is the operator for taking the integer part of a division.

And in block (2) of the flowchart in Figure 8a, the patch region extracted at each iteration is determined based on the principle:

$$\mathrm{ImgPatch}(Y, X) = i(y_s : y_e, x_s : x_e) \tag{6}$$

with

$$Starting\ point: \begin{cases} x_s = (X-1) \times s_x + 1 \\ y_s = (Y-1) \times s_y + 1 \end{cases} \tag{7}$$

$$Ending\ point: \begin{cases} x_e = x_s + p_x - 1 \\ y_e = y_s + p_y - 1 \end{cases}, \tag{8}$$

where

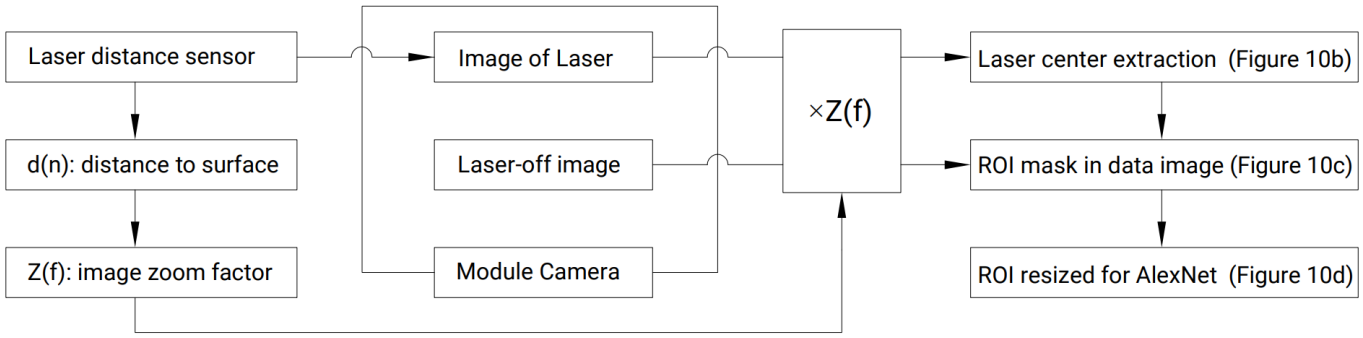

Figure 9. Data synchronization and image pre-processing for prediction.

- x_s, x_e are horizontal coordinates of the starting point and the ending point,
- y_s, y_e are vertical coordinates of the starting point and the ending point,
- X is the order of the inner loop,
- Y is the order of the outer loop,
- $i(y_s{:}\, y_e, x_s{:}\, x_e)$ is the operator for extracting a 2-D image patch based on the coordinates of the top-left point (starting point) and the bottom-right point (ending point).

2.5. AlexNet training parameter

The weights of the AlexNet model are learned during training based on data characteristics and are also influenced by predefined training parameters, which are selected according to the data properties and specific task requirements. Table 3 lists the model training parameters and their values.

According to the recommended training parameters, the dataset is split into two parts: 90 % for training and 10 % for validation. Before training begins, the data is shuffled to prevent the model from learning any unintended patterns based on data order, thereby enhancing generalization capability.

Several key training parameters were carefully selected to optimize the performance of the AlexNet model. The initial learning rate was set at 0.0003, which is relatively low compared to commonly used values (e.g., 0.01 or 0.001). This choice was based on the fine microstructural details present in the surface images. A smaller learning rate helps achieve stable convergence and prevents overshooting during weight updates. Preliminary trials using a higher rate of 0.001 resulted in fluctuating loss curves and unstable convergence, particularly when combined with data augmentation. In contrast, reducing the rate to 0.0003 led to smoother, more consistent training and improved validation accuracy.

The mini-batch size was empirically set at 10 after testing other values. Larger batches tended to reduce generalization due to the high variability in the dataset, while smaller ones introduced excessive gradient noise and reduced Graphics Processing Unit (GPU) efficiency. A batch size of 10 offered a good balance, retaining the stochastic nature of updates to escape local minima while remaining computationally efficient on mid-range hardware.

The number of training epochs was fixed at 10, based on close observation of training dynamics. The model typically stabilized after 7 epochs, with signs of overfitting observed beyond 10. Thus, setting the epoch count at 10 provided sufficient iterations for learning without compromising generalization, especially when combined with data augmentation and regular validation.

Finally, the Stochastic Gradient Descent (SGD) optimization algorithm with momentum was employed to accelerate convergence and improve training efficiency. All training parameters—learning rate, batch size, number of epochs, and optimizer—were selected not only to maximize classification accuracy but also to align with hardware constraints and ensure computational efficiency. This careful tuning helped the model achieve optimal performance while avoiding overfitting or extended training times.

Table 3. AlexNet training parameter.

No.	Parameter	Value or mode
1	Training dataset	0.9
2	Data shuffling frequency	After each epoch
3	Initial learning rate	0.0003
4	Mini-batch size	10
5	Maximum epochs	10
6	Optimization algorithm	SGD with momentum
7	Validation dataset	Augmented validation dataset

2.6. Data synchronization and image processing for prediction

Unlike conventional object recognition tasks, surface structure classification involves datasets where inter-class differences are subtle. In such cases, inconsistencies in input data, particularly in image scale (due to varying shooting distances) and resolution (pixel dimensions), can lead to reduced prediction accuracy. Therefore, aligning input images with those used during CNN training is essential for maintaining generalization when processing new data.

The overall synchronization process is illustrated in Figure 9. It includes two main pre-processing steps: (1) calculating the zoom factor using distance data from a laser sensor, and (2) resizing the Region of Interest (ROI) to 227 × 227 pixels to match the model's input requirements.

The laser sensor generates a light spot, as shown in Figure 10a, which marks both the reference distance and the ROI centre. However, this light spot consists of a cluster of high-intensity pixels, making it difficult to pinpoint the exact centre. To address this, a Gaussian filter is applied to soften the beam's edges. As shown in Figure 10b, this filter blurs the peripheral pixels, while preserving those near the centre, effectively narrowing the area where the centre is determined. The pixel with the highest intensity shifted toward the true centre is then extracted as the ROI centre.

The formula for the Gaussian function in two-dimensional image space is defined in a simple symmetric form:

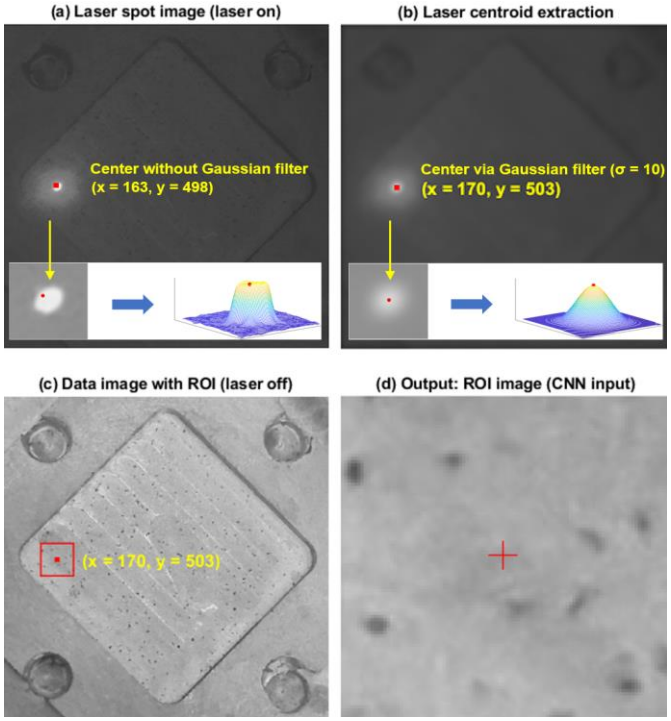

Figure 10. The process of extracting the ROI based on the laser centre point to be used as input data for CNN.

$$G(x,y) = \frac{1}{2\,\pi\,\sigma^2} e^{-\frac{x^2+y^2}{2\,\sigma^2}}, \tag{9}$$

where

- x and y are independent variables (pixel), representing coordinates in a two-dimensional plane,
- $G(x,y)$ is the value of the Gaussian function at spatial coordinates (x,y),
- σ is the standard deviation, a parameter that controls the spread or width of the distribution.

The factor $1/(2\,\pi\,\sigma^2)$ acts as a normalization factor, ensuring that the integral of $G(x,y)$ over the entire xy-plane equals one. The function reaches its maximum value at the origin $(0,0)$ decreases exponentially as distance from the origin increases.

This process ensures that the ROI images are consistent in size and scale with the image patches used during training,

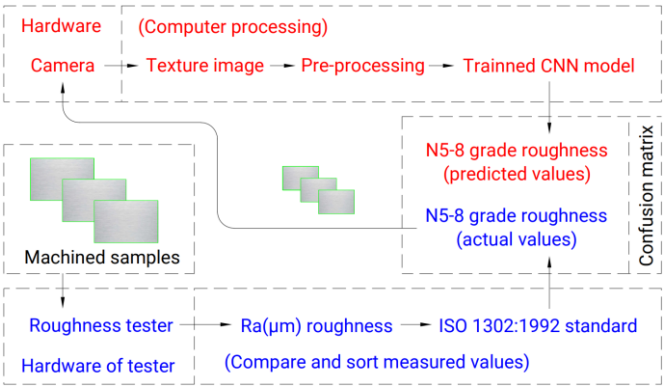

Figure 11. Framework for evaluating the performance of the model.

preserving the model's classification accuracy across varying input conditions.

3. HARDWARE AND EXPERIMENTAL SETUPS

Following the procedure described in Figure 11, the samples were independently assessed using the Mitutoyo SJ-210 roughness measurement system (Figure 13d). The diagram in Figure 12 shows the main layout and connections between hardware blocks. Subsequently, samples corresponding to levels N5–N8 (according to the ISO 1302:1992 classification standard) were selected and re-evaluated using the classification model (Figure 13b). The predicted results from the classification model were then compared with the outcomes obtained from the Mitutoyo SJ-210 device in a confusion matrix.

Figure 13a and Figure 13c show the hardware layout in the actual system. The accuracy of image acquisition for surface roughness classification is significantly influenced by image resolution and lighting uniformity.

High-resolution images provide more detailed texture information, which is essential for capturing subtle variations in surface patterns. Insufficient resolution may lead to the loss of critical microstructural features, thereby limiting the model's ability to distinguish between adjacent roughness levels.

Lighting uniformity during image capture is equally important. Uneven lighting can produce shadows or reflections that alter the perceived surface characteristics and potentially mislead the feature extraction process of convolutional neural networks. Ensuring consistent and diffused illumination helps

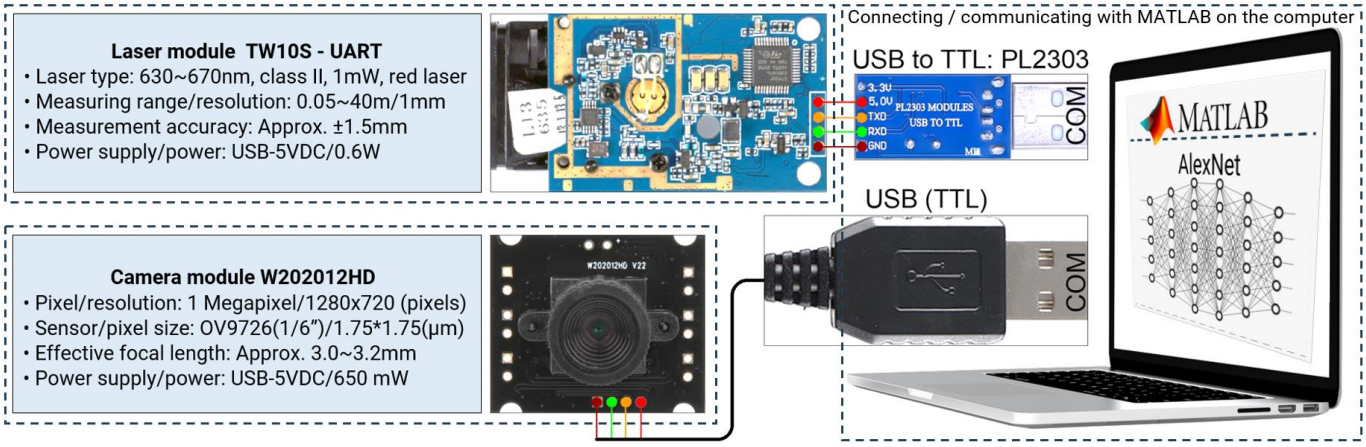

Figure 12. Specifications and diagram of hardware connection setup (camera, laser module measuring distance) with computer.

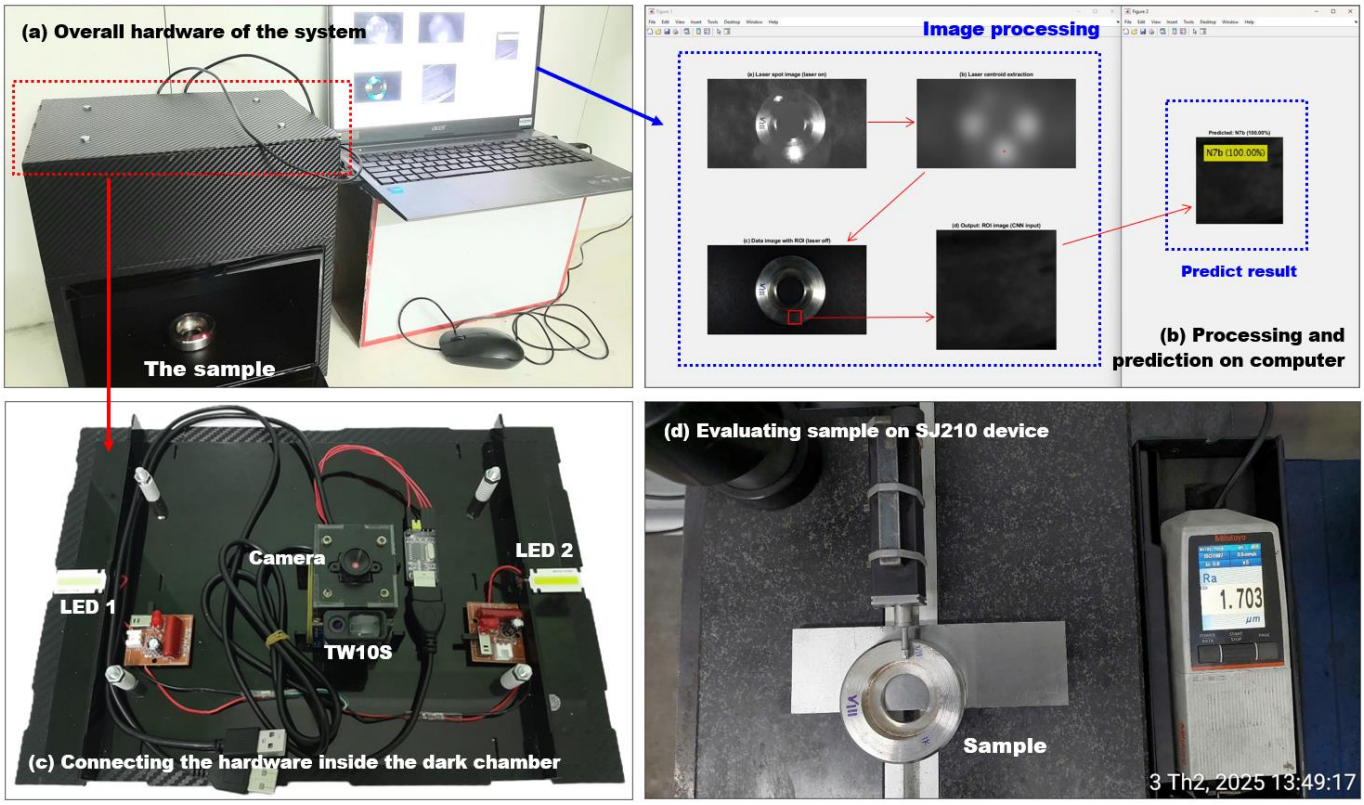

Figure 13. (a) Overall hardware of the system, (b) image processing and prediction on the computer, (c) connecting the hardware inside the dark chamber, and (d) evaluating the sample on the Mitutoyo SJ-210 device.

preserve the true texture properties of the surface and reduces intra-class variability caused by lighting-induced noise.

To minimize environmental variation and enhance image quality, active lighting control and environmental shielding measures were implemented. A controlled setup was used, including a dark chamber (lightproof enclosure) to isolate the sample from ambient light fluctuations.

Additionally, active lighting using two LED chips [a white light source combined with diffusing panels, powered by a 1.5 W supply via a USB port (3.3-5) V] was employed to provide stable

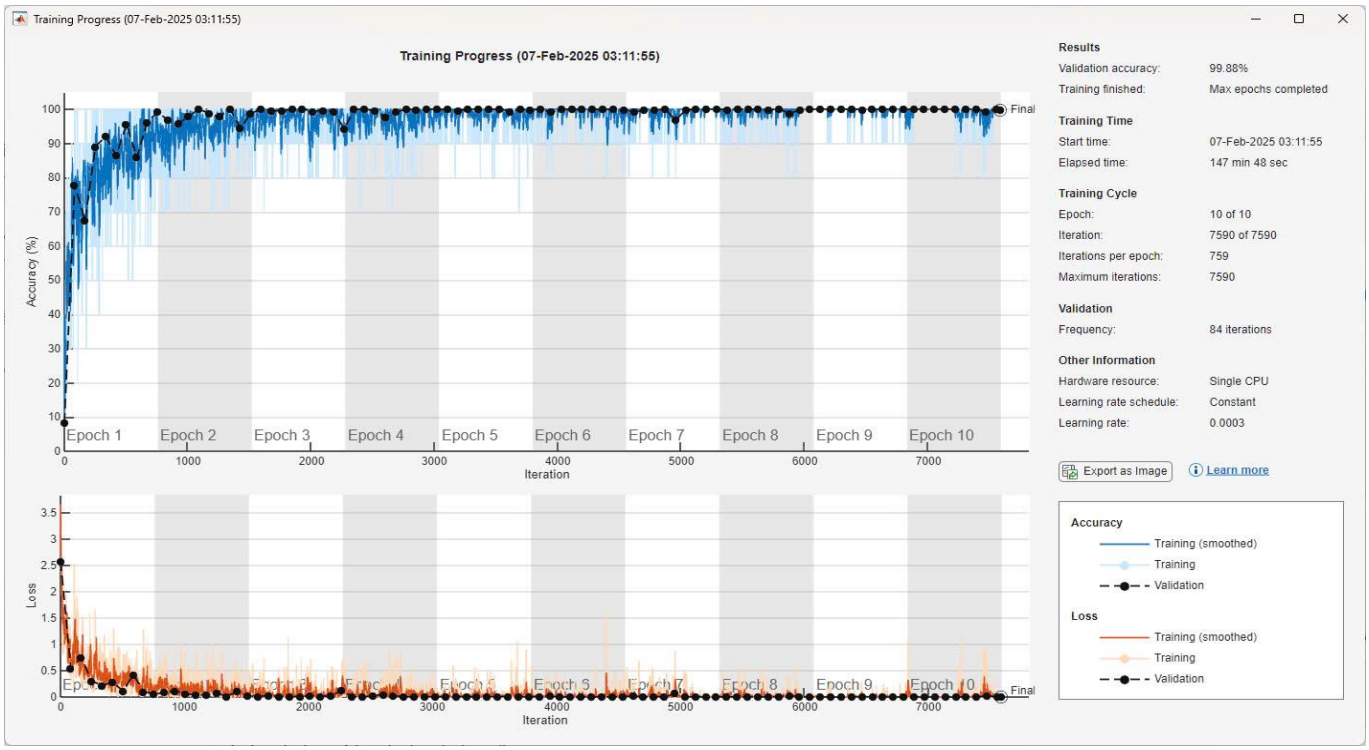

Figure 14. Model training process and classification results on the validation data set.

Table 4. The number of samples obtained in each class.

N-grade	Turning (a)	Horizontal milling (b)	Vertical milling (c)
N5	528	644	672
N6	792	748	756
N7	660	728	700
N8	882	648	630

and uniform illumination across both the imaging space and the sample surface. These measures significantly reduced noise caused by environmental light interference, improving the repeatability and consistency of image acquisition.

4. RESULTS AND DISCUSSIONS

4.1. Model training

Figure 14 shows the training progress of the AlexNet model on the collected dataset (Table 4) with the support of MATLAB 2024b: the accuracy plot (top) and the loss plot (bottom) tracking the training process. Training accuracy increases rapidly in the initial epochs and stabilizes above 95 % after 3-4 epochs. Validation accuracy remains consistently high (99 %) after the third epoch. Training loss (orange and light orange lines) decreases quickly in the early epochs and approaches zero as training progresses.

The validation loss (represented by the black dotted line) remains very low and shows minimal fluctuation after the third epoch, which suggests that the model generalizes well to unseen data and does not exhibit significant signs of overfitting.

The final validation accuracy reached 99.88 %, indicating strong performance on the test dataset. The rapid convergence suggests that the training data was well-matched to the task. No significant overfitting was observed, as both validation and training accuracies remained consistently high and stable. However, it is crucial to clearly distinguish between validation accuracy and real-world performance.

Validation accuracy reflects the model's effectiveness when evaluated on a dataset drawn from the same distribution as the training set. In real-world scenarios, the model is exposed to a broader diversity of input data, often resulting in lower actual performance, a phenomenon commonly referred to as the deployment gap.

This gap can be attributed to several factors, including limited diversity in the training dataset (in terms of materials, machining methods, and varying lighting conditions), discrepancies between training and deployment environments, and various sources of noise, such as vibration, dust, and surface reflections. The model shows high accuracy, highlighting its effectiveness for surface roughness classification. Future work should include more diverse, real-world machining data to enhance robustness in practical applications.

4.2. Real-world performance

Table 5, Table 6, and Table 7 present the confusion matrices of the model in this study, which is constructed from experimental measurement data. In this matrix, the vertical axis represents the roughness grades obtained using the SJ-210 device and classified according to ISO 1302:1992, while the horizontal axis represents the grades re-evaluated by the classification model.

Overall, it can be observed that the main diagonal exhibits low dispersion, with the elements on the main diagonal having significantly higher values compared to the off-diagonal elements. These are initial indications of a well-performing classification model.

Table 5. Confusion matrix M(r, c), r = row, c = column, in the case of turning (a)

Actual classes	Predicted classes			
	N5	N6	N7	N8
N5	9	1	0	0
N6	1	7	2	0
N7	0	1	8	1
N8	0	0	3	7

Table 6. Confusion matrix M(r, c), r = row, c = column, in horizontal milling (b)

Actual classes	Predicted classes			
	N5	N6	N7	N8
N5	10	0	0	0
N6	1	8	1	0
N7	0	2	8	0
N8	0	0	1	9

Table 7. Confusion matrix M(r, c), r = row, c = column, in vertical milling (c)

Actual classes	Predicted classes			
	N5	N6	N7	N8
N5	10	0	0	0
N6	0	8	2	0
N7	0	1	9	0
N8	0	0	0	10

The performance evaluation of a multi-class model (without data imbalance) in each class is conducted by using the precision (accuracy), recall, and F_1 (micro) score [37]:

$$F_1score_{\text{micro}} = 2 \cdot \frac{Precision_{\text{micro}} \cdot Recall_{\text{micro}}}{Precision_{\text{micro}} + Recall_{\text{micro}}}, \tag{10}$$

with

$$Precision_{\text{micro}} = \frac{TP_n}{TP_{\text{n}} + FP_{\text{n}}} \tag{11}$$

and

$$Recall_{\text{micro}} = \frac{TP_n}{TP_{\text{n}} + FN_{\text{n}}}, \tag{12}$$

where

- the true positives in class n:

$$TP_{\text{n}} = M(r,c) \text{ with } r = c \tag{13}$$

- the total false positives in class n:

$$FP_{\text{n}} = \sum_{r=1,\ r\neq c}^{n} M\,(r,c) \tag{14}$$

- the total false negatives in class n:

$$FN_{\text{n}} = \sum_{c=1,\ c\neq r}^{n} M\,(r,c)\,. \tag{15}$$

Table 8, Table 9, and Table 10 present class-wise performance metrics - precision, recall, and F1-score for different machining conditions: turning, horizontal milling, and

Table 8. $FN_n, FP_n, TP_n, Recall_n, Precision_n, F_1(n)$ in the case of turning (a)

Class	N5	N6	N7	N8
FN_n	1	3	2	3
FP_n,	1	2	5	1
TP_n	9	7	8	7
$Recall_n$	0.90	0.70	0.80	0.70
$Precision_n$	0.90	0.78	0.62	0.88
$F_1(n)$	0.90	0.74	0.70	0.78

Table 9. $FN_n, FP_n, TP_n, Recall_n, Precision_n, F_1(n)$ in horizontal milling (b)

Class	N5	N6	N7	N8
FN_n	0	2	2	1
FP_n,	1	2	2	0
TP_n	10	8	8	9
$Recall_n$	1.00	0.80	0.80	0.90
$Precision_n$	0.91	0.80	0.80	1.00
$F_1(n)$	0.95	0.80	0.80	0.95

Table 10. $FN_n, FP_n, TP_n, Recall_n, Precision_n, F_1(n)$ in vertical milling (c)

Class	N5	N6	N7	N8
FN_n	0	2	1	0
FP_n,	0	1	2	0
TP_n	10	8	9	10
$Recall_n$	1.00	0.80	0.90	1.00
$Precision_n$	1.00	0.89	0.82	1.00
$F_1(n)$	1.00	0.84	0.86	1.00

vertical milling, respectively. The results reveal noticeable variations in classification performance across different surface roughness classes (N5–N8) and machining types.

In the turning scenario (Table 8), class N5 exhibits the most balanced and reliable classification performance, with precision, recall, and F1-score all reaching 0.90. However, performance significantly deteriorates for class N7, which shows the lowest precision (0.62), indicating a high number of false positives (misclassified samples assigned to N7). Class N6 also presents a relatively low recall (0.70). This reflects difficulty in correctly identifying instances from this class. These observations suggest that the feature distribution overlap among N6 and N7 may have impaired the model's ability to draw precise decision boundaries in turning images.

The horizontal milling results (Table 9) show overall a slightly higher performance. Notably, classes N5 and N8 achieve high F1-scores of 0.95, with class N8 reaching a perfect precision of 1.00. Classes N6 and N7 demonstrate more balanced results ($\boldsymbol{Precision}$ = 0.80, $\boldsymbol{Recall}$ = 0.80, $\boldsymbol{F1}$ = 0.80), implying improved separability of features under this machining condition.

The best performance is observed in the vertical milling scenario (Table 10), where classes N5 and N8 achieve a perfect classification with precision, recall, and F1 score all equal to 1.00. Classes N6 and N7 also show substantial improvement with F1 scores of 0.86 and 0.84, respectively, a significant increase compared to turning. The precision and recall values of N6 and N7 also reached quite high thresholds, from 0.80 to 0.9.

In summary, the classification performance is highest under vertical milling, followed by horizontal milling, and lowest under turning. The N6 and N7 surface classes only slightly differ in surface roughness, which results in surface images with no clearly visible distinctions. These differences may stem from the misclassification between adjacents, which is often caused by the high similarity in image features. Consequently, feature distribution overlap may occur when extracted features from the images have a distribution that is too close, leading to overlap, making it challenging for the model to establish clear decision boundaries between adjacent classes [38]. Additionally, lighting noise or surface reflections can obscure the subtle differences between adjacent classes such as N6 and N7.

The significant difference in overall performance in actual testing (85.83 %) compared to the performance on the validation set (99.88 %) indicates some issues that need to be improved in the training dataset and hardware quality (camera). However, this is still a good result, as it demonstrates a certain level of reliability of the model in actual predictions and can be improved with a training dataset containing a larger and more diverse sample set. Besides, fine-tuning image processing or feature extraction can mitigate these issues, while potential improvements, such as oversampling [39] or label refinement [40], have been shown to be effective in most cases and do not cause overfitting in CNNs like AlexNet.

5. CONCLUSION

This study proposed a non-contact method for evaluating the surface roughness of machined parts using convolutional neural networks and image processing techniques. The results indicated that the AlexNet-based model successfully classified surface roughness grades with an accuracy of 85.83 % under experimental conditions, validating its feasibility for practical applications.

This research offers both theoretical and practical contributions to the metal machining industry. By applying image processing and deep learning for surface roughness evaluation, it enhances measurement accuracy and speed, reduces reliance on costly manual methods, and enables automation in smart production lines. The approach is adaptable across various materials and industries, such as automotive, aerospace, electronics, and precision engineering. Additionally, it supports timely production decisions by providing accurate data, improving product quality, and reducing defects. Overall, this work advances Industry 4.0 by promoting automated, intelligent quality control in manufacturing.

Despite the promising results, this model still has some limitations that need to be overcome. Accordingly, the model's performance is affected by the dataset's diversity. The dataset lacks diversity, which may contribute to performance drops in real-world testing. To further improve the model's accuracy and robustness, future work should focus on expanding the dataset by incorporating a wider range of machining methods, materials, and roughness grades, which could enhance the generalization of the model.

The confusion matrix indicates that misclassifications tend to occur between adjacent surface roughness grades, suggesting overlap in their feature representations. Enhancing the feature extraction process or incorporating additional training data may improve class separability and overall classification performance.

The camera resolution and lighting conditions impact classification accuracy, particularly for glossy or fine surface details. Enhancing the image acquisition system, particularly by using higher-resolution cameras, could improve the detection of finer surface details, making the system more applicable to precision manufacturing environments.

In future work, it will be necessary to consider and evaluate advanced architectures, such as VGG, ResNet, DenseNet (densely connected convolutional networks), EfficientNet, or ViT (vision transformers), and to compare their accuracy, generalization capability, processing speed, and computational resource requirements.

REFERENCES

[1] A. Pramanik, R. Dixit, S. Chattopadhyaya, M. S. Uddin, Y. Dong, A. K. Basak, G. Littlefair, Fatigue life of machined components, Advances in Manufacturing, 5(1) (2017), pp. 59-76. DOI: 10.1007/s40436-016-0168-z

[2] D. Taylor, O. M. Clancy, The fatigue performance of machined surfaces, Fatigue and Fracture of Engineering Materials and Structures, 14(2-3) (1991), pp. 329-336. DOI: 10.1111/j.1460-2695.1991.tb00662.x

[3] M. Field, J. F. Kahles, W. P. Koster, Surface finish and surface integrity, in: ASM International eBooks, (1989), ISBN 978-1-62708-188-7**,** pp. 19-36. DOI: 10.31399/asm.hb.v16.a0002119

[4] M. Vishnoi, P. Kumar, Q. Murtaza, Surface texturing techniques to enhance tribological performance: a review, Surfaces and Interfaces, 27 (2021), p. 101463. DOI: 10.1016/j.surfin.2021.101463

[5] F. Hashimoto, S. N. Melkote, R. Singh, R. Kalil, Effect of finishing methods on surface characteristics and performance of precision components in rolling/sliding contact, International Journal of Machining and Machinability of Materials, 6(1-2) (2009), p. 3. DOI: 10.1504/ijmmm.2009.026923

[6] ANSI/ASME B46.1: 1995 - Surface Texture, Surface Roughness, Waviness and Lay.

[7] B. Bhushan, Surface roughness analysis and measurement techniques, in: Mechanics & materials science, The mechanics and materials science series. CRC Press, (2000), ISBN 9780429126727, pp. 79-150. DOI: 10.1201/9780849377877.ch2

[8] T. R. Thomas, Characterization of surface roughness, Precision Engineering, 3(2) (1981), pp. 97-104 DOI: 10.1016/0141-6359(81)90043-x

[9] ISO 4287:1997 Geometrical Product Specifications (GPS) - Indication of surface texture: Profile method - Term, definitions and surface texture parameters - Section 4: Surface profile parameter definitions. Online [Accessed 18 June 2025] http://resource.npl.co.uk/softgauges/pdf/Specification.pdf

[10] ISO 4287:1997 Geometrical Product Specifications (GPS) - Surface texture: Profile method - Nominal characteristics of contact (stylus) instruments - Section 3: Definitions, and Section 4: Nominal values for instrument characteristics. Online [Accessed 18 June 2025] http://resource.npl.co.uk/softgauges/pdf/Specification.pdf

[11] D. J. Whitehouse, Handbook of Surface Metrology. CRC press, (2023), ISBN 978-0750300391. DOI: 10.1201/9780203752609

[12] M. Krystek, A fast gauss filtering algorithm for roughness measurements, Precision Engineering, 19(2-3) (1996), pp. 198-200. DOI: 10.1016/s0141-6359(96)00025-6

[13] K. Palova, T. Kelemenova, M. Kelemen, Measurement procedures for evaluating the surface roughness of machined parts, Applied Sciences, 13(16) (2023), p. 9385. DOI: 10.3390/app13169385

[14] ISO 12179: 2021 Geometrical product specifications (GPS) - Profile method - Calibration of contact (stylus) instruments. Online [Accessed 18 June 2025] https://www.iso.org/obp/ui/#iso:std:iso:12179:ed-2:v1:en

[15] C. J. L. Perez, Surface roughness modelling considering uncertainty in measurements, International Journal of Production Research, 40(10) (2002), pp. 2245-2268. DOI: 10.1080/00207540210125489

[16] M. A. Davinci, N. L. Parthasarathi, U. Borah, S. K. Albert, Effect of the tracing speed and span on roughness parameters determined by stylus type equipment, Measurement, 48 (2013), pp. 368-377. DOI: 10.1016/j.measurement.2013.11.023

[17] S. A. Whitehead, A. C. Shearer, D. C. Watts, N. H. F. Wilson, Comparison of two stylus methods for measuring surface texture, Dental Materials, 15(2) (1999), pp. 79-86. DOI: 10.1016/s0109-5641(99)00017-2

[18] E. S. Gadelmawla, Estimation of surface roughness for turning operations using image texture features, Proceeding of the Institution of Mechanical Engineers Part B Journal of Engineering Manufacture, 225(8) (2011), pp. 1281-1292. DOI: 10.1177/2041297510393643

[19] X. Cui, J. Zhao, C. Jia, Y. Zhou, Surface roughness and chip formation in high-speed face milling AISI H13 steel, The International Journal of Advanced Manufacturing Technology, 61(1-4) (2011). DOI: 10.1007/s00170-011-3684-9

[20] E. S. Gadelmawla, A. E. Eladawi, O. B. Abouelatta, I. M. Elewa, Investigation of the cutting conditions in milling operations using image texture features, Proceeding of the Institution of Mechanical Engineers Part B Journal of Engineering Manufacture, 222(11) (2008), pp. 1395-1404. DOI: 10.1243/09544054jem1173

[21] J. P. Davim, V. N. Gaitonde, S. R. Karnik, Investigations into the effect of cutting conditions on surface roughness in turning of free machining steel by ANN models, Journal of Materials Processing Technology, 205(1-3) (2008), pp. 16-23. DOI: 10.1016/j.jmatprotec.2007.11.082

[22] W. Rawat, Z. Wang, Deep Convolutional Neural Networks for Image Classification: A Comprehensive review, Neural Computation, 29(9) (2017), pp. 2352-2449. DOI: 10.1162/neco_a_00990

[23] T. Batu, H. G. Lemu, H. Shimels, Application of artificial intelligence for surface roughness prediction of additively manufactured components, Materials, 16(18) (2023), art. no. 6266. DOI: 10.3390/ma16186266

[24] S. Ghosh, R. Knoblauch, M. E. Mansori, C. Corleto, Towards AI driven surface roughness evaluation in manufacturing: a prospective study, Journal of Intelligent Manufacturing, (2024), pp. 1-30. DOI: 10.1007/s10845-024-02493-1

[25] A. P. Rifai, H. Aoyama, N. H. Tho, S. Z. M. Dawal, N. A. Masruroh, Evaluation of turned and milled surfaces roughness using convolutional neural network, Measurement, 161 (2020), p. 107860. DOI: 10.1016/j.measurement.2020.107860

[26] A. Giusti, M. Dotta, U. Maradia, M. Boccadoro, L. M. Gambardella, A. Nasciuti, Image-based Measurement of Material Roughness using Machine Learning Techniques, Procedia CIRP, 95 (2020), pp. 377-382. DOI: 10.1016/j.procir.2020.02.292

[27] J. Su, H. Yi, L. Ling, A. Shu, E. Lu, Y. Jiao, S. Wang, Multi object surface roughness grade detection based on Faster R-CNN, Measurement Science and Technology, 34(1) (2022), p. 015012. DOI: 10.1088/1361-6501/ac900b

[28] Y. Shi, B. Li, L. Li, T. Liu, X. Du, X. Wei, Automatic non-contact grinding surface roughness measurement based on multi-focused sequence images and CNN, Measurement Science and Technology, 35(3) (2023), p. 035029. DOI: 10.1088/1361-6501/ad1804

[29] C. Shang, Z. Lie-ping, K. A. Gepreel, H. Yi, Surface roughness measurement using microscopic vision and deep learning, Frontiers in Physics, 12 (2024), p. 1444266. DOI: 10.3389/fphy.2024.1444266

[30] A. Arif, P.G. Rao, K. Prasad, A novel hybrid Bayesian-optimized CNN–SVM deep learning model for real-time surface roughness

classification and prediction based on in-process machined surface image analysis, Int J Interact Des Manuf, (2025), pp. 1-18.
DOI: 10.1007/s12008-025-02261-2

[31] J. Meng, X. Yuan, G. Wang, X. Li, E. Zhao, J. Li, H. Fang, B. Li, C. Li, D. Zhao, H. Zhao, L. Cheng, J. Zhang, Automatic surface roughness recognition system under different manufacturing processes based on deep learning, Measurement, 253(A) (2025), p. 117473.
DOI: 10.1016/j.measurement.2025.117473

[32] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6) (2017), pp. 84-90.
DOI: 10.1145/3065386

[33] ISO 1302:1992 Geometrical Product Specifications (GPS) - Indication of surface texture in technical product documentation - Section I.3: Content of annex C of ISO 1302:1992. Online [Accessed 18 June 2025]
https://www.sharifcadcam.ir/uploaded/8ed0498b-0ddb-4d36-8914-03459fa70e1c.pdf

[34] ANSI/ASME B 46.1: 2009 Surface Texture, Surface Roughness, Waviness and Lay - Nonmandatory appendix B: Control and production of surface texture. Online [Accessed 18 June 2025]
https://www.asme.org/getmedia/71644d10-54c2-4c5d-9ab8-33ba0ba03650/28833.pdf

[35] Simonyan, K, and A Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proc. of 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 7-9 2015, pp. 1-14.
DOI: 10.48550/arXiv.1409.1556

[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016, pp. 770-778.
DOI: 10.48550/arXiv.1512.03385

[37] K. Takahashi, K. Yamamoto, A. Kuchiba, T. Koyama, Confidence interval for micro-averaged F1 and macro-averaged F1 scores, Applied Intelligence, 52(5) (2021), pp. 4961-4972.
DOI: 10.1007/s10489-021-02635-5

[38] P. Vuttipittayamongkol, E. Elyan, A. Petrovski, On the class overlap problem in imbalanced data classification, Knowledge-Based Systems, 212 (2021), pp. 106631.
DOI: 10.1016/j.knosys.2020.106631

[39] M. Buda, A. Maki, M. A. Mazurowski, A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks, Neural Networks 106 (2018), pp. 249-259.
DOI: 10.1016/j.neunet.2018.07.011

[40] H. Bagherinezhad, M. Horton, M. Rastegari, A. Farhadi, Label Refinery: Improving ImageNet Classification through Label Progression, Computer Vision and Pattern Recognition, Cornell University, (2018).
DOI: 10.48550/arXiv.1805.02641