Machine Vision and Applications (2021) 32:21
https://doi.org/10.1007/s00138-020-01142-w

ORIGINAL PAPER


# Measurement and inspection of electrical discharge machined steel surfaces using deep neural networks

Jamal Saeedi[1] · Matteo Dotta[2] · Andrea Galli[2] · Adriano Nasciuti[2] · Umang Maradia[3] · Marco Boccadoro[3] · Luca Maria Gambardella[1] · Alessandro Giusti[1]

Received: 20 April 2020 / Revised: 17 August 2020 / Accepted: 12 October 2020


**Abstract**
We propose an industrial measurement and inspection system for steel workpieces eroded by electrical discharge machining, which uses deep neural networks for surface roughness estimation and defect detection. Specifically, a convolutional neural network (CNN) is used as a regressor in order to obtain steel surface roughness and a CNN based on spatial pooling pyramid is applied for defect classification. In addition, a new method for the region of interest selection based on morphological reconstruction and mean shift filtering is proposed for defect detection and localization. The regressor and classifier based on deep neural networks proposed here outperform state-of-the-art methods using handcrafted feature extraction. We achieve a mean absolute percentage error of 7.32% on roughness estimation; on defect detection, our approach yields an accuracy of 97.26% and an area under the ROC curve metric of 99.09%.

**Keywords** Electrical discharge machining · Convolutional neural networks · Spatial pooling pyramid · Morphological reconstruction · Mean shift filtering

## 1 Introduction

The research presented here uses machine learning for measurement and inspection of steel products processed by a die-sinking electrical discharge machining (EDM) device [1]. Specifically, deep learning and image-processing techniques have been investigated, allowing for the cost-efficient integration of machine vision equipment in the EDM machine tools in order to measure steel surface roughness and inspect it for possible defects.

Die-sinking EDM comprises an electrode and a workpiece which is immersed in an insulating liquid such as oil or similar dielectric fluids. The electrode and workpiece are then connected to an appropriate pulsed voltage power supply, which generates an electrical potential between these two parts. When the electrode moves toward the workpiece, a dielectric breakdown happens between the electrodes (in the insulating liquid), which results in a plasma channel formation. EDM has been applied by numerous industries in their manufacturing developments because of its capability to easily erode hard materials, like hardened steel and tungsten carbide, and create accurate and unique shapes. The most common applications for EDM are die making, mold making, and small hole drilling [1].

Roughness measurement for die-sinking EDM steel is usually performed either with a contact profilometer or optically, utilizing interferometry or laser scanning confocal microscopy. The fact that no comparable standard exists for optical instruments still makes a contact profilometer essential in every workshop today, at least until optical measurements will reliably deliver comparable results. However, measurements using a contact profilometer is only suitable for pointwise measurements and can damage the surface because it contacts the material surface [2]; moreover, this technique is difficult to integrate into an EDM machine because of the size of the measurement device, entailing significant operational limitations with considerable risk of collision with the workpiece to be inspected. On the contrary,

✉ Jamal Saeedi
jamal.saeedi@idsia.ch

✉ Alessandro Giusti
alessandrog@idsia.ch

1 Dalle Molle Institute for Artificial Intelligence (IDSIA USI-SUPSI), Lugano, Switzerland

2 Institute for Mechanical Engineering and Material Technology (MEMTi SUPSI), Lugano, Switzerland

3 Agie Charmilles SA, Losone, Switzerland

Published online: 21 November 2020

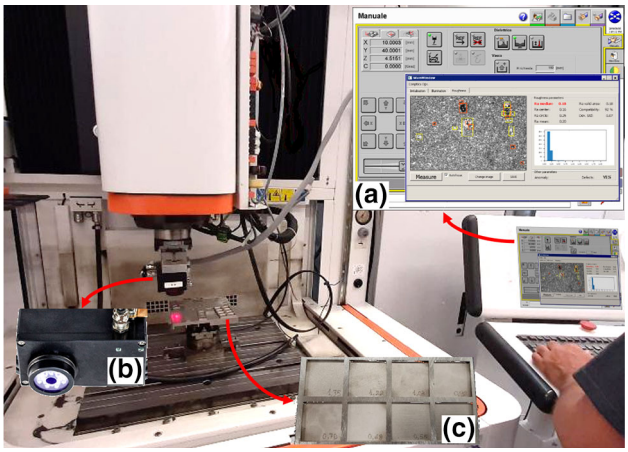


**Fig. 1** The automation system for measurement and inspection, Die-Sinking machine AgieCharmilles Form 200 by GF Machining Solutions **a** integrated GUI, **b** image acquisition system, and **c** workpiece

non-contact optical methods can be more easily integrated into the machining process.

Without an in-process measurement, if the surface roughness does not match the requirements after the process has concluded, there is no way of correcting the error. This is due to the fact that the exact repositioning of workpiece and electrode, and recreating the microscopic gap conditions which are required to resume machining, is in many cases impossible. Depending on the type, size, and complexity of the workpiece, this can yield very high costs. Therefore, a vision-based technique, which delivers comparable results to the contact profilometer has a high potential for the EDM industry.

Zero defect manufacturing (ZDM) is now a key concept as industry 4.0 is minimizing human involvement in production processes. As a result, publications related to defect inspection have grown rapidly in the past decade. However, publications related to the visual inspection of EDM steel are rare and mostly related to the geometrical defects [3]: in contrast, defects on EDM steel surfaces have a complex visual appearance, often with unclear definition, size, and shape, and are challenging to discriminate from the underlying background [4].

Motivated by these requirements, this paper presents a machine-integrated, inexpensive optical measurement and inspection system with two main goals: measuring surface roughness values comparable to the results originating from contact profilometer, and inspecting the steel surface for defects. The system (shown in Fig. 1) relies on a standard machine vision camera mounted, instead of an electrode, on the chuck of a die-sinking EDM machine.

The paper is organized as follows. After a review of related work (Sect. 1.1), the main contribution of the paper is presented in Sect. 2: a method for EDM steel surface measurement and inspection, which involves anomaly detection, roughness estimation, and defect detection and localization. Section 3 demonstrates experimental results and discussion which describes experimental setup, dataset, evaluation metrics, roughness estimation, and defect detection results. Finally, conclusions and future works are discussed in Sect. 4.

### 1.1 Related work

Surface roughness measurement methods can be either contact (e.g., profilometers) or non-contact (e.g., optical). Contact-based methods are simple and widespread, but difficult to integrate into the machining process, and may potentially damage the steel surface [2]. In order to solve these problems and ensure the workpiece quality, non-contact methods have been proposed as early as four decades ago [5].

Some non-contact methods rely on ad hoc sensors. Bradley et al. [6] adopt a fiber optics sensor for surface roughness measurement, in which the surface topography is observed through phase changes of the incident and reflected light on the steel surface. Sato et al. [7] proposed a method based on utilizing a scanning electron microscope for surface roughness estimation. They demonstrated that the profile of a surface could be obtained by processing back-scattered electron signals, which correspond to the surface inclination along with the electron beam scanning direction; then, the profile of the surface roughness can be derived by integrating the intensity of the back-scattered electron signal. Infrared scatter-meters were also adopted for roughness measurements of engineering surfaces [8]. Khan et al [9] proposed a method based on artificial neural networks (ANN) to model Titanium steel surface roughness machined by EDM. The machining parameters including peak current, pulse-on time, pulse-off time, and servo-voltage are used as inputs to a multilayer perceptron (MLP) for roughness prediction.

Instead of using specialized sensors, a cost-efficient alternative is to rely on standard machine vision cameras or polarized microscopes for non-contact roughness measurement using image-processing and machine-learning techniques. Several related works use this approach to measure the characteristics of a surface [10]. A similar approach adopts machine learning using input images acquired by polarized microscopes [11], or use specialized techniques to reconstruct the 3D topography for roughness estimation [12].

Various classifications have been made for the vision-based methods that can be distinguished into four general categories including spectral, statistical, time series analysis, and learning-based approaches. The most common technique for texture analysis in the spectral-based methods is Fourier transform. In an earlier work in [13] a 2-D fast Fourier transform (2D-FFT) of the digitized surface image is used as a measurement parameter of the steel surface. Since the Fourier transform does not have the necessary ability to detect local

irregularities (ignore local deviations), a number of methods have been developed based on multi-resolution image analysis, e.g., wavelet transform [10], and curvelet transform [14] for roughness estimation.

Statistical methods often adopted for texture analysis rely on features such as Local Binary Patterns [45], Haralick features of the gray level co-occurrence matrix [15], and histograms [16]; time series analysis methods predict the surface roughness by taking into account the dynamic characteristics of images such as fractal [17, 18] and the largest Lyapunov exponent [10].

Learning-based approaches are the last category and can be divided into two sub-categories consisting of traditional feature extraction and deep learning-based methods. The conventional method starts with the feature extraction, followed by classic machine learning or ANN approaches. All of the spectral, statistical, and time series analysis techniques can be used for feature extraction in this sub-category. In [19], an MLP neural network was used to model and predict the optical roughness values using the 2D-FFT of the surface image as feature. Morala et al. [20] proposed a method to predict the surface roughness using MLP along with the wavelet transform of the surface image as a feature. The methodology is based on the extraction of texture features from surface images in the frequency domain using wavelet transform and a roughness classification using a MLP neural network. Tsai et al. [21] assessed the surface roughness of machined parts produced by the shaping and milling processes and extracted features of surface roughness in the spatial frequency domain using the 2D-FFT. These roughness features were taken as input to ANN to estimate the surface roughness. Another similar approach has been applied for carbon steel and aluminum alloy after being processed by milling [11], in which an MLP is used to predict roughness based on a binary input image.

However, the end-to-end approach combines feature extraction and the roughness estimation into one whole using deep learning neural networks framework in which features are automatically extracted through the learning of training sets. The deep learning-based approach is simple and reaches high detection accuracy. Sun et al. in [2] proposed a method based on CNN for milled metal surface roughness estimation, which involves texture skew correction, image filtering, and applying neural network for roughness estimation. Our approach operates on images from die-sinking EDM, which have a significantly different, challenging appearance; we follow a similar approach and adopt convolutional neural networks (CNN) architectures; we also implement and compare the results with feature-extraction-based approaches.

Defect inspection usually involves two steps including defect detection and defect classification [4]. Defect detection approaches rely on statistical, filtering, model-based, and machine-learning techniques to identify defects and then using supervised or unsupervised classification methods to assign the class label of the detected defects (for further information see [4]). In this work, we focus on defect detection, and compare different approaches and classification algorithms. To this end, two approaches are common in the literature: patch-based methods, which densely classify all possible patches in the input image, and ROI-based methods, which only operate on a subset of ROIs. Patch-based approaches are generally computationally expensive, since a huge number of possible positions and sizes must be considered for the patch used as an input for classification. In this paper, we adopt an alternative approach, which relies on image-processing techniques for determining a limited number of ROIs; on each ROI, we then apply binary classification for defect detection.

Generally, ROI extraction is performed using edge detection, thresholding, or clustering for segmenting the region or detect the edges from an image. The purpose of edge detection is to identify the areas of an image where a large change in intensity occurs. Edge detection is usually performed utilizing local linear-gradient operators such as Sobel, zero cross edge detection, and Canny and Roberts cross edge detection methods [22]. Thresholding is another method applied for segmentation in numerous image-processing applications. An object having homogeneous intensity and a background with a different intensity level could easily be differentiated with this approach. Global thresholding, adaptive thresholding, histogram equalization, and Otsu thresholding are different types of thresholding techniques [23]. The segmented image resulted from the thresholding methods provide more information for subsequent processing tasks. Clustering is another method that tries to find the relationships among patterns of the dataset by organizing them into groups or clusters [24]. K-means clustering and fuzzy C-means clustering are the widely used clustering techniques.

Because of the complex appearance of EDM images, our approach implements a fusion of morphological reconstruction and mean shift filtering is proposed for ROI extraction. Having the defect candidates, the next step is to apply a binary classification to determine which defect candidates can actually be considered as a defect.

As mentioned previously, the main objectives of this research project are applying an image-based regression for steel surface roughness estimation and a binary classification for defect detection. Commonly, there are two methods which are applied for image-based regression and classification tasks: conventional feature extraction and deep learning-based methods. The conventional method starts with the feature extraction, followed by a classification (or regression) and a result output. The feature extraction process needs artificial feature designing, which sometimes could be a tedious and complicated task. However, the end-to-end approach combines feature extraction and the classification

(or regression) processes into one whole using deep learning neural networks framework in which features are automatically extracted through the learning of training sets [25, 26]. The deep learning-based approach is simple and reaches high detection accuracy. However, its main disadvantage is the need for a large number of training images which should cover sufficient defect types; otherwise, detection results would not ideal. This problem is usually treated by introducing augmentation techniques to increase the training dataset. Because of the excellent performance of convolutional networks based on deep learning in the field of image processing and machine learning, it will be investigated in this study for the roughness estimation through regression and the defect classification.

## 2 Proposed method

After analyzing the steel surface images taken by the camera system, we now describe the system design (see Fig. 2) that fulfills two objectives: (a) roughness estimation and (b) defect detection and localization. The first step is anomaly region detection, which involves detecting the out-of-focus and poorly illuminated area in the image; this typically occurs when the image acquisition system reaches the edge of the workpiece being inspected. Since we are using an image-based measurement for roughness estimation, the anomalous region needs to be ignored. This is performed by first identifying the anomalous area and then applying impainting by selecting random parts of the image from the valid region of the steel surface.

Defect detection is solved in two steps. In the first step, we apply image-processing techniques to detect defect candidates in the image, and make sure the procedure has high sensitivity while accepting low specificity—i.e., the algorithm should not miss (almost) any defect, but might detect several false defects. In the second step, we use heuristics and machine-learning approaches to determine which defect candidates can actually be considered as a defect. In the following sub-sections, the proposed methods for anomaly detection, roughness estimation, and defect detection and localization will be explained in detail.

### 2.1 Anomaly detection

By studying the dataset, it is observed that there is an issue related to the out-of-focus and poorly illuminated area in the image (see Fig. 3). Since a low depth-of-field camera is used for image acquisition, it can be assumed that the anomalous area in the image is the out-of-focus region. Here, an algorithm based on Laplacian transform is used to identify the out-of-focus and poorly illuminated area as the anomalous region [27, 28]. The Laplacian operator is a marginal point detection operator that is independent of the edge direction, and it is a second-order differential operator. Considering $f(x, y)$ as a continuous two elements function, the Laplacian operation can be defined as follows:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{1}$$

where $x$ and $y$ are the standard Cartesian coordinates of the $xy$-plane, and $\partial^2$ is the second derivative.

For digital images, the Laplacian operation can be simplified as convolving with a kernel:

$$g(x, y) = \sum_{i=-L}^{L} \sum_{j=-K}^{K} f(x - i, y - j) H(i, j) \tag{2}$$

The Laplacian highlights regions of an image containing rapid intensity changes. Therefore, if an image contains high variance regions then there is a widespread of responses applying the Laplacian operator, which is associated with the normal or in-focus region. But if the variance is very low, then there will be a little spread of responses for the Laplacian operator, indicating that there are very few edges in the region, which is associated with the out-of-focus region.

Anomaly detection is started by applying the Laplacian operator on the image. It is followed by a global thresholding using a predefined threshold tuned on a large dataset. After thresholding, the morphological image filling is performed to fill the small holes. Since there is only one dominant in-focus region in the image, the region with the maximum area will be selected as a valid region (see Fig. 3). The anomalous region is impainted with part of the steel surface from the valid region in the image to avoid complications in the roughness estimation step (see Fig. 3). It should be mentioned that the defect detection module will only be applied to the valid region in the steel surface image.

### 2.2 Roughness estimation

Roughness is traditionally evaluated by moving a contact stylus over a linear path on the surface and filtering the resulting profile to obtain several quantitative surface roughness parameters as described in ISO 4287 [29]. The most common parameter used in workshops to qualify the surface roughness is the so-called average roughness (Ra) value, defined by ISO 4287:1997. The Ra is the arithmetic average deviation from the mean line in the machined zone, and it can be calculated by using the following:

$$R_a = \frac{1}{L} \int_0^L |Z(x)| dx \tag{3}$$

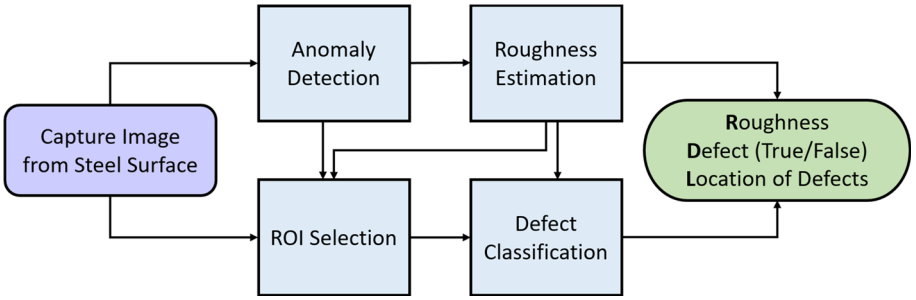


**Fig. 2** Block diagram of the proposed image-based measurement and inspection system

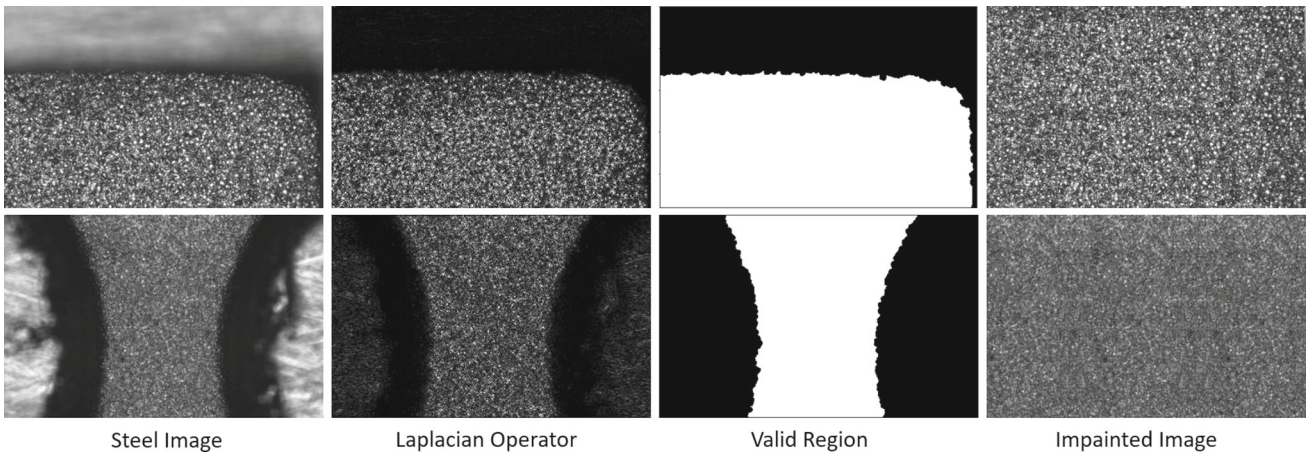


**Fig. 3** Anomaly detection and impainting for roughness estimation task

where $L$ is the evaluation length and $Z$ is the roughness profile height as a function of $x$.

Figure 4 shows a number of samples of the EDM steel surfaces with different roughnesses. Here, the roughness estimation task is modeled as an image-based regression problem using the data provided by the image acquisition system integrated with the EDM machine. Having the dataset containing the images from the steel surface and measured Ra values (Ra between 0.14 and 3.6 μm), the task is to train a deep CNN to estimate the Ra for the new unseen images from the EDM steel surface [30]. A CNN is a special kind of multi-layer neural network, designed to recognize visual patterns directly from images with minimal pre-processing. CNNs are usually adopted for classification tasks (i.e., in problems where the target variable is categorical), but can be also applied for regression tasks (i.e., to estimate a scalar quantity): to achieve that, we remove the fully connected softmax classifier layer typically used for classification, replacing it with a fully connected layer with a single node with a rectified linear unit (ReLU) activation function. The model is trained using a regression loss (e.g., mean squared error). A similar application which is using CNN for image-based regression is vision-based real estate price estimation [31]. This is known as a supervised regression problem and is structured in two sequential steps. In the first step, the CNN is trained using a training dataset, a collection of many training instances in which each instance is a pair of one input image and the corresponding Ra value of the surface measured by a contact profilometer. This step is time consuming and performed only once, off-line, on a powerful computer. In the second step, the trained CNN is copied to the EDM machine and used to predict the Ra value for new unseen input images from new EDM steel surfaces.

Some of the CNN architectures of ImageNet large-scale visual recognition challenge (ILSVRC) top competitors consisting of LeNet-5 [32], AlexNet [33] and VGGNet-16 [34] have been modified here for the regression task instead of classification, i.e., the last layer has been changed to ReLU activation function instead of softmax for classification. We have used simple CNN architectures to limit the amount of weights and allow fast inference on CPU-only memory-limited deployment machines. There are other alternatives for the CNN architectures, e.g., MobileNet [35], ResNet [36], and Xception [37], which can be applied for the regression task here at the expense of slower training and inference times. The three networks considered for the regression task here consist of 5, 8, and 16 layers, characterized by a sequence of alternating convolutional and Max-Pooling layers; which is inspired by the visual cortex of the human brain. In addition, AlexNet [33] and VGGNet-16 [34] incorporate dropout and batch normalization functionalities. Dropout functionality allows the CNN network to randomly set a fraction rate of

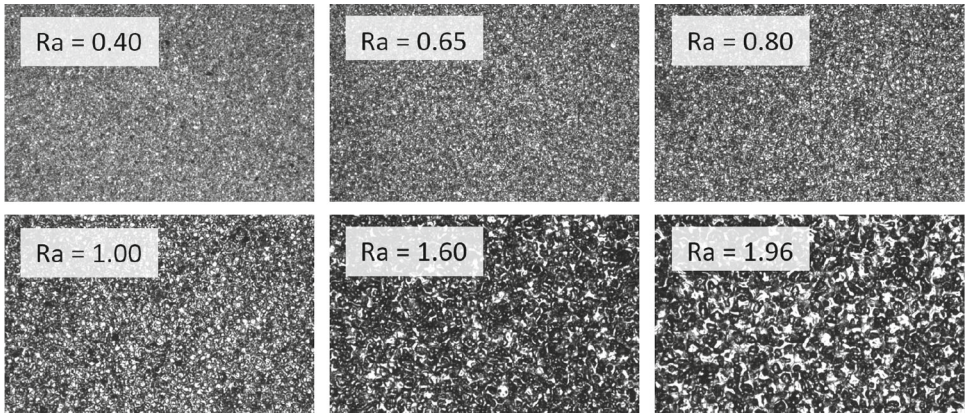


**Fig. 4** Image samples showing the EDM steel surfaces with different roughnesses

input units to 0 at each weight update during the training step, which helps prevent overfitting. A batch normalization layer is also used to transform inputs so that they are standardized, i.e., they will have a mean of zero and a standard deviation of one, which helps reduce the sensitivity to the initial starting weights, allows much higher learning rates for increasing the speed at which networks train.

Since the size of the steel surface image is large ($740 \times 480 \times 1$) and contains much redundant information for the task of estimating the roughness, the image is divided into four quadrants and then used for the training ($370 \times 240 \times 1$). The estimated roughness for the image can then be obtained using the mean of the results of these four quadrants. The input image and the roughness values are rescaled between the range [0, 1] for the training. The training process consists in determining a value for each of the network parameters, in such a way that, for each image in the training set, the predicted output value is as close as possible to the corresponding desired output (i.e., the true Ra value of the surface visible in the image given as input). The loss is the mean absolute percentage error between the predicted and desired outputs, and the network is trained using stochastic gradient descent (batch size of 128) using the ADAM optimizer [38], with an initial learning rate of 0.0001. By separating 10% of the training set as a validation set, the model that gives the best result on the validation set over 300 epochs will be stored. The trained network will then be used for roughness estimation (Ra value) for the test images.

Other alternatives to CNN for regression tasks using state-of-the-art methods are also implemented and compared and will be presented in the experimental results section.

## 2.3 Defect detection and localization

This section presents the proposed approach for defect detection and localization task. Generally, there are three different situations when looking for defects in steel products [4]. Defects can be simply detected in most situations by using standard image-processing techniques such as thresholding and segmentation. A slightly more difficult situation is the case that the defects are not very clear in terms of their size and shape; however, it can still be recognized from the underlying background. The most complicated case is a defect in which its definition, size, and shape are not clear, and it is difficult to distinguish the defect from the underlying background. The defects that are visible in the EDM steel surface are an example of the third situation because of the complex background due to different roughnesses.

By analyzing the defected steel images, it is observed that there are two types of defects named as black and white defects available in the dataset as shown in Fig. 5. The bright defects could be the area with a roughness higher than the rest of the steel surface (i.e., area where the electrical discharge machine worked the most in the roughing process), or the area damaged by the mechanical scratch. The dark defects are because of the burn (arc-spot) or hydrocarbon-dielectric decomposition product that is dried on the surface, or dirt due to material in the tank (hairs of rags or particles of materials) [39].

The proposed approach for ROI selection is a combination of different image-processing techniques which is shown in Fig. 6. The key idea behind the method that is used for ROI selection is to divide *peak* and *valleys* regions in order to better identify the defected regions. In this way, the defected regions become more differentiated from the surrounding area. The first step toward ROI selection is image denoising using the bilateral filtering, which is a non-linear, edge-preserving, and noise-reducing smoothing filter for images. In the bilateral filtering, the intensity of each pixel is replaced with a weighted average of intensity values from neighboring pixels. These weights are obtained based on a Gaussian distribution. It is followed by a grayscale morphological reconstruction (MR) step to divide the denoised image into two images named as *peaks* and *valleys,* which are representing bright and black defects respectively.

Grayscale reconstruction is one of the important techniques provided by mathematical morphology [40], which is mainly applied for several filtering, segmentation, and fea-

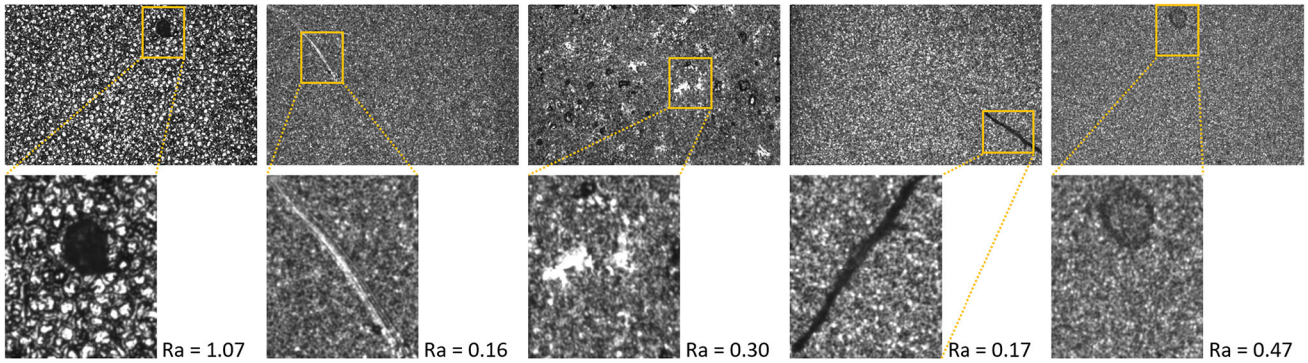


**Fig. 5** Image samples showing the EDM steel surfaces with different types of defects and roughnesses

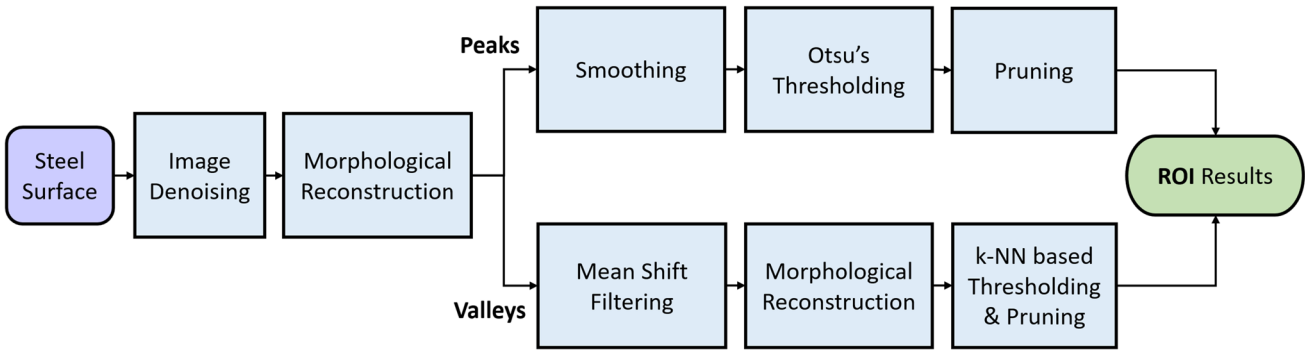


**Fig. 6** Block diagram of the proposed ROI selection method

ture extraction tasks. If we assume in an image the x and y axes represent pixel positions and the z-axis represents the intensity of each pixel, there would be a three-dimensional map in which the intensity values represent elevations, as in a topographical map. In topographical terms, the areas with low and high intensities in an image, *valleys* and *peaks*, are important morphological features because they can often spot important image objects.

In order to find *valleys* regions in the image, morphological reconstruction is started by the erosion operation. An image named as seed is initialized to the local maximum values of the original image. It should be mentioned that along the borders of seed image, the original values of the image are used. Erosion expands the *minimal* values of the seed image until it encounters a mask image. These border pixels will be the starting points for the erosion process. The erosion is then limited by setting the mask to the values of the original image. The eroding started from the edges removes holes, since the holes by definition are surrounded by pixels of brighter value. Finally, the dark regions are isolated by subtracting the reconstructed image from the original image named as *valleys*.

Alternatively, bright spots in an image (named as *peaks* image) can be found using morphological reconstruction by the dilation operation. Dilation operation acts as the inverse of erosion and it expands the *maximal* values of the seed image until it encounters a mask image. Since this is an inverse operation, the seed image is initialized to the local minimum values of the original image instead of the maximum. The remainder of the process is the same. The *peaks* and *valleys* images are then separately being processed to obtain bright and dark defect candidates as shown in Fig. 6. The *valleys* image is then smoothed by the mean shift filtering algorithm [41].

The idea behind the method that is used to enhance the image is the fact that human knowledge about images of the physical world is that they are spatially smooth, in the sense that neighboring pixels are more likely to belong to the same object (class) than to different ones. Mean shift filtering method is a clustering algorithm which is commonly used in computer vision and image-processing applications. First, for each pixel of an image having a spatial location and a particular color, a set of neighboring pixels within a spatial radius and a defined color distance is determined. Then, for this set of neighboring pixels, the new spatial center (spatial mean) and the new color mean value will be calculated. These calculated mean values will be used as the new center for the next iteration. The described procedure will be iterated until the spatial and the color (or grayscale) mean stops changing. At the end of the iteration, the final mean color will be assigned to the starting position of that iteration. One of the advan-

tages of mean shift filtering is the edge-preserving capability compared to other alternative smoothing techniques.

The next step is to apply the MR once again to find *valleys* in the smoothed image and to ignore homogeneous areas in the thresholding step. Because of the complexity of the image's background due to different roughnesses and also different illuminations, a method based on k-nearest neighbors ($k$-NN) is applied for image thresholding. First, a small set of diverse images have been selected for manually selecting the thresholds for making the train-set. Then a $k$-NN regressor (with $k = 1$) is trained using mean, standard deviation, and roughness of the images as input and integer threshold as output. Finally, the threshold will be selected using trained $k$-NN. After thresholding and pruning small-sized objects in the binary image, the isolated objects in the binary image will be considered as dark defect candidates.

Another conceptually simple algorithm applied on the *peaks* image to find ROI for bright defect candidates. Similar to the *valleys* image, the peaks image is smoothed using a Gaussian filtering approach for a lighter smoothing compared to the mean shift filtering applied for *valleys* image. It is followed by image thresholding using Otsu's method [23]. Pruning based on the object's size in the binary map is applied to remove unlikely defect candidates. The threshold for pruning is adaptively obtained based on the image roughness, the lower is the roughness the lower will be the threshold. It should be noted that the defect candidates that have overlap with edge region (which is obtained by anomaly detection) will be neglected for the classification step. Some of the important steps of the ROI selection algorithm are visually demonstrated by applying on a sample defected steel image in Fig. 7.

Having the ROI candidates, the final task is to classify them as defects or not. Generally, there are two ways to deal with images with different sizes using deep learning-based classification. The first approach is resizing the images into fixed-size representation preserving the aspect ratio or not. The second one is to apply a feature selection method after convolutional layers and before dense layers to generate a fixed-length feature vector. The latter shown promising results for image classification tasks [42]. Therefore, a similar strategy is applied here for different ROIs to provide a binary classification.

He et al. proposed a pooling strategy called spatial pyramid pooling (SPP) layer in order to provide a fixed-length vector before dense layers for CNN-based classification [43]. The convolutional layers accept arbitrary input sizes; however, they produce outputs of variable sizes. A fixed-length feature vector is a requirement for the classifiers like support vector machine (SVM) or fully connected layers in the case of CNN. This kind of fixed-length vectors can be created using the spatial pyramid pooling [44], which can preserve spatial information through pooling in local spatial bins. Since the spatial bins have proportional sizes with respect to the image size, therefore the number of bins will be fixed regardless of the image size. Standard CNNs use sliding window pooling in contrast to SPP [43], where the number of sliding windows will be dependent on the input size. In order to adopt the CNN for images of arbitrary sizes, the last pooling layer should be replaced with a SPP *layer*. In each spatial bin, it will pool the responses of each filter. The outputs of the SPP will be $kM$-dimensional vectors with the number of bins denoted as $M$ ($k$ is the number of filters in the last convolutional layer). The fixed-dimensional vector will be used as the input to the fully connected layer.

Inspired by [42], a new architecture consisting of three convolutional layers followed by max pooling, one convolutional layer followed by SPP and three dense layers is used here for the task of binary classification (see Fig. 8). The activation function for convolutional layers is ReLU, and the sigmoid function is selected for the dense layers. In addition, the dropout function is used with a rate of 15% in dense layers to avoid overfitting problems.

The network accepts images with different sizes as input, however, since two pooling layers are applied (one max pooling and one SPP), the input image width and length should be equal or larger than eight pixels. Therefore, slightly larger windows around the defects are extracted as ROI in order to consider not only the size limit but also the neighboring pixels. Since the Ra is being available for each image being inspected for defect detection, thanks to roughness estimation using CNN, it can be used as an auxiliary input for the binary classification task. Considering the fact that there is a limited number of defected images having a limited number of roughnesses in the training set, it is decided to give the network only a number between one to five from very low to very high roughnesses, in order to limit the chance of overfitting. The roughness class is added to the network after the SPP layer (see Fig. 8).

The input image is rescaled between the range [0, 1], and the categorical cross-entropy or log loss is used as the loss function for the training. The proposed network is trained using stochastic gradient descent using the ADAM optimizer [38], with an initial learning rate of 0.00003. By separating 10% of the training set as a validation set, we save the model that gives the best result on the validation set over 300 epochs, (which requires about 24 h on a 24-CPU workstation equipped with 4 T K80 GPUs). Since we have images with different sizes, batch processing is not possible, therefore, the network training is time consuming; however, it is an offline processing task and will be performed only once. The trained networks will be used for binary classification task given the ROIs for defect candidates as input.

Other alternatives to CNN-based binary classification using state-of-the-art feature extraction methods are also

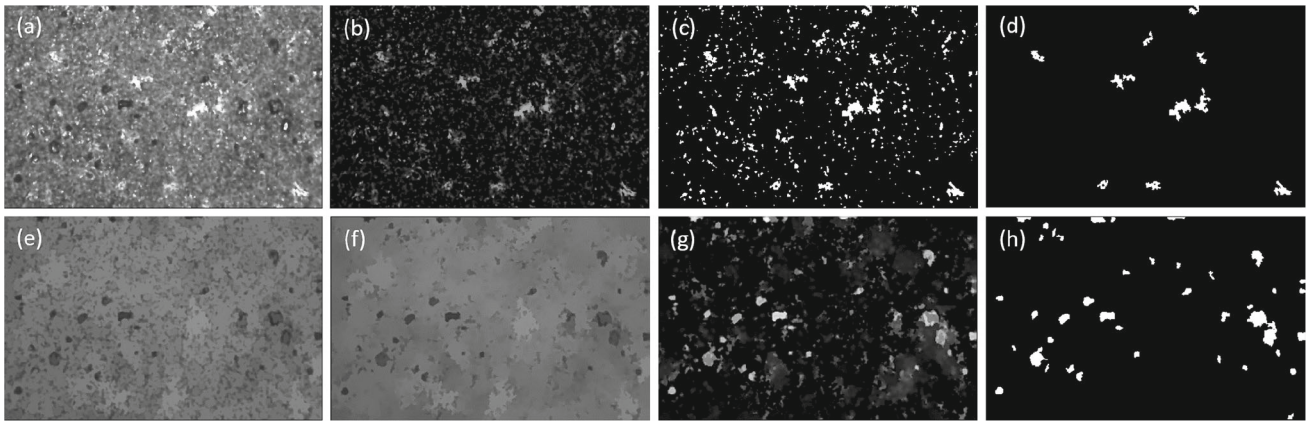


**Fig. 7** Visual representation of proposed ROI selection method, **a** steel image, **b** *peaks* image, **c** Otsu thresholding, **d** ROI result for bright defects after pruning, **e** *valleys* image, **f** mean shift filtering, **g** morphological reconstruction, and **h** ROI result for dark defects after *k*-NN-based thresholding and pruning

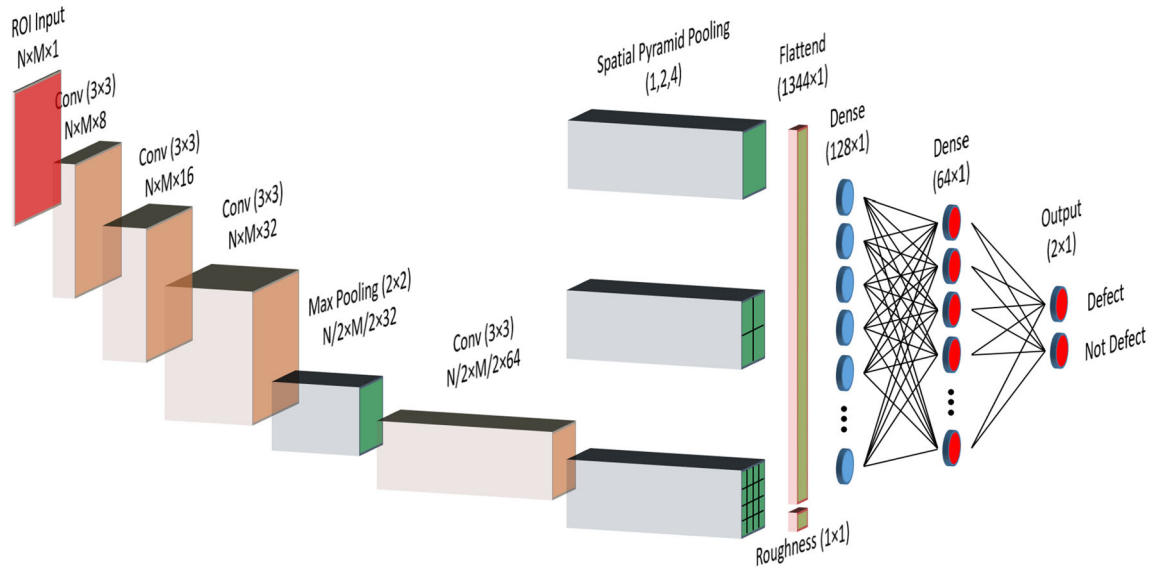


**Fig. 8** Proposed architecture used for defect classification; the input ROI (top left) is processed by a sequence of convolutions (Conv), max-pooling and spatial pyramid pooling layers; three fully connected (dense) layers yield two output neurons, whose activations are mapped to the probabilities of being defect and not

implemented and compared and will be presented in the next section.

## 3 Experimental results and discussion

In this section, the results of the proposed methods for roughness estimation and defect detection tasks will be presented. In addition, it will be discussed how to collect data and evaluate the proposed framework as well as several state-of-the-art approaches.

The classic feature extraction-based approach is also considered here for comparison. Specifically, the state-of-the-art image features consisting of histogram (H), Hu moments (Hu), Haralick texture features (HT), local binary patterns (LBP) [45], histogram of oriented gradients (HOG) [46], and Haar-like features (Haar) [47] are used for both regression and classification tasks.

For the regression task, the state-of-the-art models including SVM with radial basis function kernel [48], and extreme gradient boosting (EGB) with Tweedie regression [49] are implemented and compared. Roughness estimation results using the CNN-based network consisting of MobileNet [35], ResNet [36], and Xception [37] have been also implemented and compared. Since 2D-FFT feature has been applied in earlier works [13, 19, 21], a method based on 2D-FFT fea-

ture extraction and a three layers MLP network has been implemented and fine-tuned in order to provide a baseline for roughness estimation results.

For the defect classification task, the state-of-the-art models consisting of the $k$-NN, decision tree (DT), SVM [48], random forests (RF) [50], and EGB [49] are implemented and compared. The CNN with LeNet-5 [33] architecture training with a resized ROI dataset ($64 \times 64$) is also implemented for comparison. Deeper networks also can be considered for the classification task, however since there may be a lot of defect candidates in a single image, it will be very time consuming. Different combinations of features are rescaled in the range of [0, 1] and applied to provide different results for the regression and classification tasks. It should be mentioned that parameter tuning is performed for different models included in the comparison in order to find the best solution for them.

### 3.1 Experimental setup

The EDM machine used for this project is built by GF Machining Solutions (see Fig. 1). The camera used to take pictures of the workpiece is custom made by Conoptica AS, a Norwegian company specialized in camera-based measurement solutions. This particular camera is able to auto-illuminate the surface of the workpiece and focus extremely close to it. Thanks to that it is possible to have a detailed image of the surface and see how much is eroded by the machine. It uses a circular light-emitting diodes (LEDs) ring light that produces uniform diffuse front illumination, which can be considered as dark field lightening [51]. In the dark field lighting method, the light source and the CCD sensor are on the same side of the steel surface. In this case, the reflection angle is not equal to the incident angle, and the line between the CCD sensor and the image of the light source is not on the same line as the reflected light; therefore, it is difficult for light to enter the CCD sensor.

### 3.2 Dataset

The dataset for the roughness estimation task was acquired on different samples, each machined from a different steel typically used to produce molds. The samples were machined on an AgieCharmilles Form 200 LTC die-sinking EDM machine. On each sample, a set of cavities with different Ra values were machined, the nominal Ra values range from 0.14 to 3.6 μm. The samples were cleaned in an ultrasonic bath prior to measurement. The roughness of the bottom surface of each cavity was measured with the line profiling method using a contact stylus surface profiler (Taylor Hobson Form Talysurf 120). Five measurements have been made automatically for each cavity. The result was an average value of Ra as defined by ISO 4287 [29]. A series of non-overlapping images covering the bottom surface of each cavity has been acquired automatically using a sensor mounted on the chuck of the EDM die-sinking machine. The field of view was approximately $2 \times 1.3$ mm with a resolution of $752 \times 480$ (approximately 2.6 um/pixel). The image was stored lossless in PNG format, without any processing. The exposure parameters were kept constant and were chosen to avoid over- or under exposure and to minimize noise. The sensor used a built-in illuminator and was not sensitive to ambient light. The resulting dataset contains 14,000 images, which is divided into 11,200 images in a training set and 2800 in a testing set (20%). The dataset has been analyzed and impainted if necessary by an anomaly detection algorithm, which is described in Sect. 2.

Due to being trained on labeled data, feature extracted by CNN is much different from the handcrafted features, which is considered data-dependent. In general, the generalization ability is considered to benefit from the large-scale labeled data taken into the process of the training model. In practice, more data lead to more discriminative information learned and less prone to overfit. Deepid [52] for face verification tasks is a good example, which introduces more data into model training to get a better result. Therefore, the augmentation techniques including horizontal and vertical flipping, and random brightness augmentation (well-suited for the task) are used to generate more training data. The brightness of the image can be augmented by either randomly darkening images or brightening images. The training set involves 100,000 images after the augmentation.

The dataset for the defect detection task was generated using 500 defected images with different roughnesses and 1200 random clean images (Ra values range from 0.14 to 3.6 μm). The proposed ROI algorithm and manual labeling are used to generate the dataset including 24,000 ROIs (20% used as test set) images with different sizes including 6000 defected and 18,000 clean samples. The augmentation techniques including horizontal and vertical flipping, random brightness augmentation, and transposing are applied to generate around 350,000 images for the training set. Since it is not possible to use batch processing for training of the CNN based on SPP, first we use the training set without augmentation to train the network and then augmentation is applied for fine-tuning the network.

### 3.3 Evaluation metrics

The evaluation metrics employed for the regression task are mean absolute error (MAE) and mean absolute percentage error (MAPE). Both of them are popular measures for evaluating the accuracy of prediction models. Equations (4) and

(5) give the definitions for these two metrics, where $p_i$ is the predicted value and $t_i$ is the true value for the $i$-th instance:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|t_i - p_i| \tag{4}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{t_i - p_i}{t_i}\right| \tag{5}$$

The evaluation metrics employed for the binary classification tasks are classification accuracy (ACC), area under the curve (AUC), and time complexity. ACC is the number of correct predictions made as a ratio of all predictions made:

$$\text{ACC} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \tag{6}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

The AUC is obtained by measuring the area under the receiver operating characteristics (ROC) curve. The ROC curve is created by plotting the true positive (TP) against the false positive (FP) at various threshold settings. The AUC represents a model's ability to discriminate between positive and negative classes. Time complexity (the amount of time it takes to run the algorithm) is also used for comparing different algorithms. The results for time complexity are averaged over 1000 runs.

### 3.4 Roughness estimation results

It should be noted that the same training and testing split are used to evaluate all the approaches. Table 1 shows the regression results for different feature combinations and models. From the results, it can be seen that the handcrafted features applied for roughness estimation reaches to 12.93% for MAPE and fails to generate comparable results to the deep learning-based method as shown in Table 2, which is 7.32%.

**Table 1** The regression results for different combinations of features and models

| Combination (#dim) | Feature | | | |
|---|---|---|---|---|
| | SVM | | EGB | |
| | MAE | MAPE | MAE | MAPE |
| $F_1$: H, Hu, HT (55) | 0.203 | 26.89 | 0.201 | 16.62 |
| LBP, $F_1$ (105) | 0.155 | 17.32 | 0.146 | 15.52 |
| HOG, $F_1$ (115) | 0.204 | 22.43 | 0.180 | 17.13 |
| Haar, $F_1$ (136) | 0.257 | 20.59 | 0.289 | 27.82 |
| LBP, HOG, $F_1$ (165) | 0.163 | 16.95 | 0.158 | 13.72 |
| LBP, Haar, $F_1$ (186) | 0.181 | 19.26 | 0.183 | 18.18 |
| HOG, Haar, $F_1$ (196) | 0.194 | 20.32 | 0.200 | 18.59 |
| All features (246) | 0.132 | 13.03 | **0.128** | **12.93** |

It can be observed from the regression results for different CNN architectures in Table 2 that the VGG-16 in this case outperformed other models for the roughness estimation task based on MAE metric. Another reason besides using a deeper network could be due to applying dropout function in the dense layers of VGG-16 to avoid overfitting and therefore leading to a better generalization compared to the LeNet-5 architecture. The experiments with the newly developed CNN architectures like MobileNet [35], ResNet [36], and Xception [37] yielded similar quantitative performances with higher inference times.

The result for the 2D-FFT feature with a three-layer fine-tuned MLP method, which is used as a baseline here, is shown a large performance gap compared to the proposed CNN-based approach. The 2D-FFT feature has been applied in earlier works [13, 19, 21] for roughness estimation. It is due to the fact that the EDM surface image does have some dominant frequencies but it is not periodic, and therefore the roughness cannot be easily captured exclusively by 2D-FFT as feature.

From the time complexity point of view, considering the large gap of the performance between MLP and CNN-based methods as shown in Table 2, the increase of time complexity from 0.0089 (seconds) for MLP to 0.0143 for CNN (LeNet-5) would be an acceptable trade-off. It is also interesting that the deep learning-based approach almost runs 12 times faster than the classic feature extraction method.

On average, the MAE and MAPE for steel surfaces in the range [0.14, 1.0] are 0.023 and 6.51%, which grow to 0.128 and 8.56% for range [1.0, 2.0]. The errors increases to 0.217 and 8.43% for Ra > 2.0 μm. The ground truth roughness values measured using line profiling method versus estimated roughnesses obtained by the deep learning-based approach in the test set are shown in Fig. 9.

### 3.5 Defect detection results

The defect classification results using state-of-the-art classifiers and deep learning-based methods are shown in Tables 3 and 4, respectively. It can be seen that unlike the regression task, the classic methods could generate comparable results to the deep learning-based methods applied here. However, the network proposed in this paper based on SPP outperformed other approaches in terms of two metrics consisting of ACC, and AUC.

A simple and effective way of boosting the performance of classification is to combine the output of several classifiers. If the errors of different classifiers have zero mean and are uncorrelated with each other, then the average error could be reduced by a factor of $M$ by simply averaging the output of the M models [53]. Here two CNN-based classifiers including

**Table 2** The regression results for different CNN architectures and the best feature combination

| Metric | 2D-FFT MLP | CNN architectures | | | | | | Best F.C |
|---|---|---|---|---|---|---|---|---|
| | | LeNet-5 | AlexNet | VGG-16 | MobileNet | Xception | ResNet | |
| MAE | 0.159 | 0.1110 | 0.1007 | **0.0960** | 0.0968 | 0.1003 | 0.1089 | 0.128 |
| MAPE | 15.63 | 7.732 | **7.324** | 7.561 | 7.930 | 7.982 | 8.164 | 12.93 |
| Time complexity (370 × 240) (s) | **0.0089** | 0.0143 | 0.0189 | 0.0256 | 0.0465 | 0.0521 | 0.0630 | 0.296 |

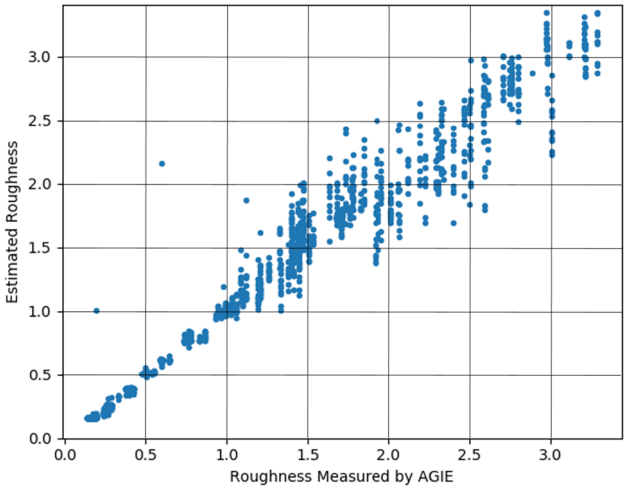


**Fig. 9** Regression result estimated vs real measured value by a contact profilometer

the proposed network and LeNet-5 trained on the resized dataset are combined and used as a hybrid method. It can be clearly seen from Table 4 that using a committee of classifiers considerably boosts the ACC up to 97.26 and AUC up to 99.09.

In terms of time complexity, the deep learning-based approaches are much faster than the classic methods as shown in Tables 4. It is important due to the fact that there are a lot of ROIs as defect candidates in an image for the classification task, e.g., if there are 20 defect candidates in an image, the hybrid method is 200 times faster than the best feature combination and model. Figures 10 and 11 demonstrate both the ROI detection and the defect classification results for sample images with different roughnesses and different defect types. The ROI selection algorithm proposed here works very well to consider all of the possible candidates for the classification step. It should be mentioned that the classification results presented here are based on the selected ROIs in the image, not the whole image individually.

In terms of error analysis, one error that can possibly happen is related to the ROI selection algorithm. It is possible that the ROI selection misses some of the defects due to the image's poor resolution and wrong threshold selection; however, based on the empirical results, it is observed that the proposed ROI selection is accurate, and very sensitive not to

**Table 3** The classification results for different combinations of features and models

| Feature Combination (#dim) | 4-NN | | DT | | SVM | | RF | | EGB | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | AUC | ACC | AUC | ACC | AUC | ACC | AUC | ACC | AUC |
| $F_1$: H, Hu, HT (55) | 95.83 | 91.58 | 71.11 | 66.80 | 95.49 | 96.62 | 96.13 | 98.12 | 96.23 | 97.55 |
| LBP, $F_1$ (89) | 96.13 | 97.06 | 84.42 | 71.92 | 94.57 | 97.16 | 96.33 | 98.71 | 96.83 | 98.72 |
| HOG, $F_1$ (119) | 94.79 | 93.86 | 89.96 | 81.63 | 95.39 | 96.74 | 96.35 | 98.70 | **96.92** | 98.76 |
| Haar, $F_1$ (136) | 94.69 | 89.80 | 73.60 | 70.58 | 94.89 | 96.13 | 96.17 | 98.28 | 96.53 | 98.65 |
| LBP, HOG, $F_1$ (153) | 94.93 | 95.36 | 85.84 | 76.46 | 93.79 | 97.21 | 96.35 | 98.80 | 96.77 | 98.77 |
| LBP, Haar, $F_1$ (170) | 95.93 | 95.97 | 85.86 | 71.84 | 93.17 | 96.92 | 95.83 | 98.38 | 96.75 | 98.73 |
| HOG, Haar, $F_1$ (200) | 93.59 | 92.34 | 88.26 | 79.23 | 95.11 | 96.67 | 96.19 | 98.36 | 96.59 | **98.82** |
| All features (234) | 93.59 | 93.66 | 86.48 | 77.40 | 93.53 | 97.18 | 96.61 | 98.55 | 96.69 | 98.69 |

**Table 4** The classification results for CNN architectures and the best feature combination

| Metric | CNN Architectures | | | Best F.C |
|---|---|---|---|---|
| | Proposed | LeNet-5 | Hybrid | |
| ACC (%) | 97.14 | 96.65 | **97.26** | 96.59 |
| AUC (%) | 98.84 | 98.52 | **99.09** | 98.82 |
| Time complexity (64 × 64) (s) | 0.0053 | **0.0013** | 0.0065 | 0.064 |

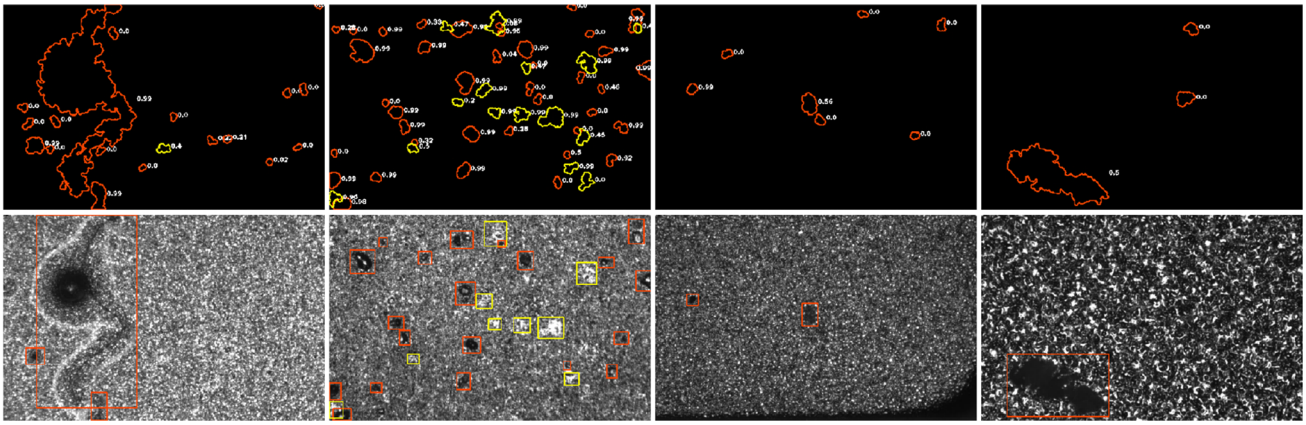

**Fig. 10** Samples of defect detection and localization results; top row: defect candidates (dark (orange) and bright (yellow)) along with probabilities of defect class, bottom row: final results

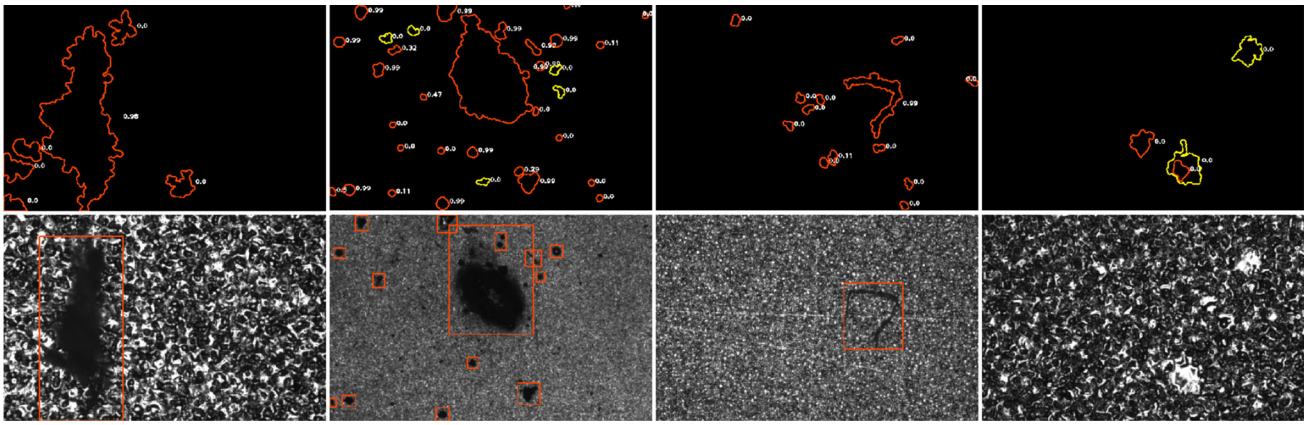

**Fig. 11** Samples of defect detection and localization results; top row: defect candidates (dark (orange) and bright (yellow)) along with probabilities of defect class, bottom row: final results

miss any defect. Another error is related to the classification error for selected ROIs, which is shown in Fig. 11 (rightmost image). This is happening because we had very few defect samples for the images with high roughness values and the network could not have generalized to correctly classify the selected ROIs in this case. For the rest of the images shown in Figs. 10 and 11, the network correctly generalized. It should be mentioned that we have applied both data augmentation and regularization (using dropout function inside the network to avoid overfitting) in order to provide better generalization capability for the proposed network for defect classification.

## 4 Conclusions and future works

We proposed a novel framework for steel surface roughness estimation and defect detection, after being processed by a die-sinking EDM machine. In particular, the proposed framework is able to perform supervised feature extraction directly from the pixel representation of the steel images using deep neural networks: this does not require the algorithm designer to hand-craft features, which is a significant advantage with visually complex inputs. In addition, a novel ROI selection algorithm is proposed here based on the morphological reconstruction and mean shift filtering for defect detection and localization. Extensive quantitative experiments on real-world data show that the CNN-based regressor and classifier we propose outperform state-of-the-art methods using feature extraction approaches: in particular, we obtain a mean average percentage error of 7.32% for roughness estimation and a defect classification accuracy of 97.26%. Our work shows a new promising application of deep neural networks for the automation of die-sinking EDM processes. For future work, we will investigate detecting additional types of defects and extend the binary classification problem proposed here to a multi-class case. Adapting the models to higher-resolution image acquisition systems can also be considered to reach better results for the roughness estimation and defect classification.

**Acknowledgements** This work has been supported by the Commission for Technology and Innovation (Project 27359.1 PFIW-IW) and the Canton of Ticino (Switzerland) through the SUPSI EDM Competence Centre. The authors would like to thank the anonymous reviewers for their helpful and constructive comments that significantly contributed to improving this paper.

## References

1. Jameson, E.C.: Electrical Discharge Machining. SME, Dearborn, Michigan (2001)
2. Sun, W., Yao, B., Chen, B., He, Y., Cao, X., Zhou, T., Liu, H.: Non-contact surface roughness estimation using 2D complex wavelet enhanced ResNet for intelligent evaluation of milled metal surface quality. Appl. Sci. **8**(381), 1–24 (2018)
3. Wang, J., Sanchez, J., Iturrioz, J., Ayesta, I.: Geometrical defect detection in the wire electrical discharge machining of fir-tree slots using deep learning techniques. Appl. Sci. **9**(90), 1–8 (2018)
4. Sun, X., Gu, J., Tang, S., Li, J.: Research progress of visual inspection technology of steel products—a review. Appl. Sci. **8**(2195), 1–25 (2018)
5. Luk, F., Huynh, V.: A vision system for in-process surface quality assessment. In: Proceedings of the Vision, SME Conference, Detroit, Michigan, pp. 12–43 (1987)
6. Bradley, C., Bohlmann, J., Kurada, S.: A fiber optic sensor for surface roughness measurement. J. Manuf. Sci. Eng. **120**, 359–367 (1998)
7. Hisyoshi, S., Masanori, O.: Surface roughness measurement by scanning electron microscope. Ann. CIRP **31**, 457–462 (1982)
8. Bjuggren, M., Krummenacher, L., Mattsson, L.: Non contact surface roughness measurement of engineering surface by total integrated infrared scattering. Precis. Eng. **20**, 33–45 (1997)
9. Khan, M.A.R., Rahman, M.M., Kadirgama, K.: Neural network modeling and analysis for surface characteristics in electrical discharge machining. Procedia Eng. **90**, 631–636 (2014)
10. Pour, M.: Determining surface roughness of machining process types using a hybrid algorithm based on time series analysis and wavelet transform. Int. J. Adv. Manuf. Technol. **97**, 2603–2619 (2018)
11. Samtaş, G.: Measurement and evaluation of surface roughness based on optic system using image processing and artificial neural network. Int. J. Adv. Manuf. Technol. **73**, 1–4 (2014)
12. Atieh, A.M., Rawashdeh, N.A., AlHazaa, A.N.: Evaluation of surface roughness by image processing of a shot-peened, TIG-welded aluminum 6061–T6 alloy: an experimental case study. Materials **11**(5), 1–18 (2018)
13. Hoy, D.E.P., Yu, F.: Surface quality assessment using computer vision methods. J. Mater. Process. Technol. **28**(1–2), 265–274 (1991)
14. Raju, R.S.U., Ramesh, R., Raju, V.R., Mohammad, S.: Curvelet transforms and flower pollination algorithm based machine vision system for roughness estimation. J. Opt. **47**(2), 243–250 (2018)
15. Zhou, L. et al: Study on brittle graphite surface roughness detection based on gray-level co-occurrence matrix. In: Proceedings of the 3rd International Conference on Mechanical, Control and Computer Engineering (2018)
16. Ghodrati, S., Kandi, S.G., Mohseni, M.: A histogram-based image processing method for visual and actual roughness prediction of sandpaper. In: Proceedings of the 6th International Congress on Color and Coatings (2015)
17. Majumdar, A., Bhushan, B.: Role of fractal geometry in roughness characterisation and contact mechanics of surface. J. Tri. ASME **112**, 205–216 (1990)
18. Saeedi, J., Faez, K., Moradi, M.H.: Hybrid fractal-wavelet method for multi-channel EEG signal compression. Circuits Syst. Signal Process. **33**(8), 2583–2604 (2014)
19. Fadare, D.A., Oni, A.O.: Development and application of a machine vision system for measurement of surface roughness. J. Eng. Appl. Sci. **4**(5), 30–37 (2009)
20. Morala-Argüello, P., Barreiro, J., Alegre, E.: A evaluation of surface roughness classes by computer vision using wavelet transform in the frequency domain. Int. J. Adv. Manuf. Technol. **59**, 213–220 (2012)
21. Tsai, D.M., Chen, J.J., Chen, J.F.: A vision system for surface roughness assessment using neural networks. Int. J. Adv. Manufac. Technol. **14**, 412–422 (1998)
22. Luo, S., Yang, J., Gao, Q., Zhou, S., Zhan, C.A.: The edge detectors suitable for retinal OCT image segmentation. J. Healthcare Eng. **3978410**, 1–13 (2017)
23. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging **13**(1), 146–165 (2004)
24. Sathya, B., Manavalan, R.: Image segmentation by clustering methods: performance analysis. Int. J. Comput. Appl. **29**(11), 27–32 (2011)
25. Yi, L., Li, G., Jiang, M.: An end-to-end steel strip surface defects recognition system based on convolutional neural networks. Steel Res. Int. **88**, 176–187 (2016)
26. Masci, J., Meier, U., Ciresan, D., Schmidhuber, J.: Steel defect classification with max-pooling convolutional neural networks. In: International Joint Conference on Neural Networks, Brisbane, QLD, Australia. IEEE, Piscataway, NJ, USA, vol. 20, pp. 1–6 (2012)
27. Saeedi, J., Faez, K.: A classification and fuzzy-based approach for digital multi-focus image fusion. Pattern Anal. Appl. **16**(3), 365–379 (2013)
28. Saeedi, J.: Image fusion in the multi-scale transforms domain using fuzzy logic and particle swarm optimization. Master dissertation, Amirkabir University of technology (2010)
29. ISO 4287 Geometrical Product Specifications (GPS)—Surface texture: Profile method—Terms, definitions and surface texture parameters (1997)
30. Boccadoro, M., Giusti, A., Gambardella, L.M.: Method for machining and inspecting of workpieces. European Patent No. EP3326749 B1 (2016)
31. You, Q., Pang, R., Cao, L., Luo, J.: Image-based appraisal of real estate properties. IEEE Trans. Multimedia **19**(12), 2751–2759 (2017)
32. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
33. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2012)
34. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2015)
35. Andrew, G.H. et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861 (2017)
36. He, K., Zhang, X., Ren, S. Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770–778 (2016)
37. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807 (2017)
38. Chollet, F.: Deep Learning with Python. Manning Publications Co, Shelter Island (2017)
39. Maradia, U., Scuderi, M., Knaak, R., Boccadoro, M., Beltrami, I., Stirnimann, J., Wegener, K.: Super-finished surfaces using Meso-micro EDM. In: Proceedings of the 17th CIRP Conference on

Electro Physical and Chemical Machining (ISEM), pp. 157–162 (2013)
40. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans. Image Process. **2**(2), 176–201 (1993)
41. Dorin, C., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
42. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
43. Grauman, K., Darrell, T.: The pyramid match kernel: discriminative classification with sets of image features. In: 10th IEEE International Conference on Computer Vision, vol. 1, pp. 1458–1465 (2005)
44. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
45. Ojala, T., Pietikinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)
46. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. Int. Conf. Comput. Vis. Pattern Recogn. **1**, 886–893 (2005)
47. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2001)
48. Cortes, C., Vapnik, V.N.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
49. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. ArXiv e-prints (2016)
50. Ho, T.K.: Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, pp. 278–282 (1995)
51. Neogi, N., Mohanta, D.K., Dutta, P.K.: Review of vision-based steel surface inspection systems. J. Image Video Proc. **50**, 1–19 (2014)
52. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. arXiv preprint arXiv:1412.1265 (2014)
53. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Berlin (2006)



**Jamal Saeedi** received his M.Sc. and Ph.D. degrees in Electronic Engineering from Amirkabir University of Tehran, Iran, in 2010 and 2015, respectively. He works in the field of signal/image processing and computer vision, specializing particularly in information fusion, pattern recognition and deep learning, road traffic monitoring systems, industrial inspection, and synthetic aperture radar imaging.

**Matteo Dotta** is a mechanical engineer employed as senior researcher and lecturer at the MEMTI institute of SUPSI.

**Andrea Galli** is a microtechnology engineer/researcher at the MEMTI institute of SUPSI. Over the years, he gained experience and specialization in mechatronics, robotics, production and automation technology, and sensor technology.

**Adriano Nasciuti** studied electrical engineering at the ETH in Zürich and went on to earn a PhD at the same institution in 1995. He continued his professional career working for an international machine tool industry in different responsibility positions inside the R&D department. Since 2012, he has been the Head of the Mechanical Engineering and Materials Technology Institute of SUPSI. In addition, he is active since 2017 in the Innovation Council of Innosuisse, the Swiss Innovation Agency. His research interests are manufacturing technologies with particular focus on the electrical discharge machining.

**Umang Maradia** received his M.Sc. and Doctorate degrees in Mechanical Engineering from ETH Zurich, Switzerland, in 2014. He is currently the head of Innovation EDM at GF Machining Solutions, Switzerland.

**Marco Boccadoro** received an MSEE from ETHZ Zurich in 1978. Then he attended advanced courses at UCLA Los Angeles in the field of microcontrollers. Later he joined Agie Charmilles as Head of Generator and Process Control Department, Product manager EDM Die Sink machines, Head of Research & Innovation, SKU Swiss school of corporate management. Since 2019, he is working as a consultant. He has held academic positions at ETHZ (lecturer of nonconventional manufacturing processes) and lecturer at Politecnico di Milano (EDM, 2018). He has authored and co-authored 18 patents and published papers for CIRP, ISEM, and in trade magazines of the industry.

**Luca Maria Gambardella** is an internationally recognized AI expert, with 35 years of experience in the field. He is a full Professor at the Faculty of Informatics of USI in Lugano and Professor at IDSIA USI-SUPSI, which he directed for 25 years. In his career, he has successfully negotiated, conducted, and delivered many AI and ML projects with Swiss and international Research Agencies and Companies. Scientifically, he is one of the pioneers of the Ant Colonies Optimization Metaheuristics. He has published more than 300 scientific articles, with h-index 70 and more than 55000 citations. He is the Co-founder and CTO, Head of Applied AI at Artificialy in Lugano, Switzerland.

**Alessandro Giusti** holds a PhD in Computer Science from Politecnico di Milano, 2008; since then, he is with the Dalle Molle Institute for Artificial Intelligence (IDSIA), where he is now a permanent Senior Researcher and Head of the robotics lab. He took part in a dozen projects in applied research focusing on deep learning applications to mobile and industrial robotics; he is the author of more than 80 peer-reviewed publications in top conferences and journals and the recipient of several awards, most of which for innovative applications of deep learning to various fields.