

Article

# Product Quality Prediction for Wire Electrical Discharge Machining with Markov Transition Fields and Convolutional Long Short-Term Memory Neural Networks

Jehn-Ruey Jiang \*  and Cheng-Tai Yen

Department of Computer Science and Information Engineering, National Central University,  
Taoyuan City 32001, Taiwan; 107522111@cc.ncu.edu.tw

\* Correspondence: jrjiang@csie.ncu.edu.tw

**Abstract:** This paper proposes a wire electrical discharge machining (WEDM) product quality prediction method, called MTF-CLSTM, to integrate the Markov transition field (MTF) and the convolutional long short-term memory (CLSTM) neural network. The proposed MTF-CLSTM method can accurately predict WEDM workpiece surface roughness right after manufacturing by collecting and analyzing static machining parameters and dynamic manufacturing conditions. The highly accurate prediction is due to the following two reasons. First, MTF can transform data into images to extract data temporal information and state transition probability information. Second, the CLSTM neural network can extract image spacial features and temporal relationship of data that are separated far apart. In short, MTF-CLSTM predicts WEDM workpiece surface roughness with the MTF model and the CLSTM neural network using static machining parameters and dynamic manufacturing conditions. MTF-CLSTM is compared with 10 related research studies in many aspects. There is only one existing method that is like MTF-CLSTM to predict WEDM workpiece surface roughness by using static machining parameters and dynamic manufacturing conditions. Experiments are conducted to evaluate MTF-CLSTM performance to show that MTF-CLSTM significantly outperforms the existing method in terms of the prediction mean absolute percentage error.

**Keywords:** Markov transition field; convolutional neural network; long short-term memory; wire electrical discharge machining; surface roughness



**Citation:** Jiang, J.-R.; Yen, C.-T. Product Quality Prediction for Wire Electrical Discharge Machining with Markov Transition Fields and Convolutional Long Short-Term Memory Neural Networks. *Appl. Sci.* **2021**, *11*, 5922. <https://doi.org/10.3390/app11135922>

Academic Editor: Giangiacomo Minak

Received: 30 May 2021  
Accepted: 22 June 2021  
Published: 25 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

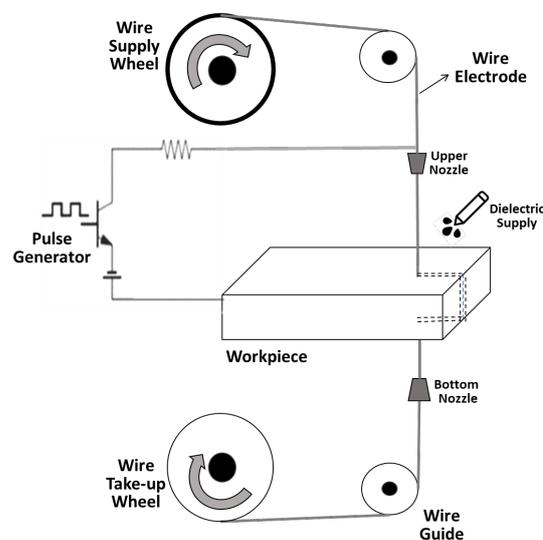
Driven by artificial intelligence (AI) and internet of things (IoT) technologies, manufacturers are paying more and more attention to smart manufacturing [1]. Predicting product quality is fundamental in smart manufacturing. For some manufactured products whose quality cannot be measured speedily or handily, it is desirable to fast and accurately predict the product quality based on static data, such as manufacturing parameters tuned before production, as well as dynamic data, such as manufacturing conditions gathered during production.

This paper focuses on product quality prediction for wire electrical discharge machining (WEDM) [2]. In practice, it focuses on predicting surface roughness  $R_a$  of the WEDM product. The surface roughness  $R_a$  is defined as the arithmetic mean of the absolute values of the profile deviations from the mean line of the roughness profile. Surface roughness has strong influence on product properties, such as friction, corrosion resistance, wear resistance, light reflection, holding lubricant, heat transfer and distribution, strength, and fatigue strength. Hence, this paper focuses on predicting WEDM product surface roughness rather than many other product qualities, such as the drum-shaped error, geometric accuracy, and so on.

WEDM is a thermo-electrical process that can produce complex 2D and 3D shapes from electrically conductive workpieces by using sparks of electrical discharges. The schematic

diagram of a WEDM machine is shown in Figure 1 [3]. The workpiece material and a wire electrode are subject to a pulse voltage, usually of tens or even hundreds of volts. However, they are separated by dielectric fluid (e.g., deionized water), so they are insulated. The wire electrode is usually made of copper, brass, or tungsten. It is wound between two spools, and travels at a constant velocity. The workpiece is moved toward the wire electrode for machining. When the workpiece is very close to the wire electrode (e.g., when the gap between them is less than a few  $\mu\text{m}$ ), the insulation is broken and a plasma channel is formed in a small area. Discharge occurs between the workpiece and the wire electrode, generating sparks that produce intense heat with temperatures of  $8000\text{ }^{\circ}\text{C}$  to  $12,000\text{ }^{\circ}\text{C}$  to melt or even vaporize workpiece material. The heat also vaporizes the dielectric fluid, causing large explosion to remove (or flush away) workpiece material debris. The pulse voltage can be used to control discharges. When the pulse is off, the discharge stops and the insulation remains. When the pulse is on again, discharges reoccur to remove workpiece materials. Repeating voltage pulse-on and pulse-off periods can thus achieve the purpose of machining materials, even of high strength and toughness.

The WEDM product quality, such as the surface roughness, is affected by many machining parameters, including the pulse-on time, pulse-off time, open voltage, gap voltage, peak current, wire tension, wire material, wire diameter, wire feed rate, servo feed rate, dielectric flushing pressure, dielectric flow rate, conductivity of dielectric fluid, workpiece height, and thermal conductivity of workpieces, etc. [3]. Due to the large number of parameters and their combinations, researchers usually fix some parameters and changed only few parameters to perform WEDM experiments for gathering data. The gathered data are then used for analyzing and modeling to optimize WEDM processes [4–14] and predict WEDM product quality [15–28].



**Figure 1.** Schematic diagram of WEDM.

A method, called MTF-CLSTM, is proposed in this paper to integrate the Markov transition field (MTF) [29] and the convolutional long short-term memory (CLSTM) neural network for WEDM product quality prediction. The MTF is used to represent dynamic WEDM manufacturing conditions as images. The images are then fed into a convolutional neural network (CNN) [30] to further extract features. Finally, the extracted features, along with static manufacturing parameters, are fed into a long short-term memory (LSTM) neural network [31] to predict the surface roughness of the WEDM product right after manufacturing. MTF-CLSTM is compared with related work [19–28] in many aspects. There is only one existing method [27] that is like MTF-CLSTM to predict WEDM workpiece surface roughness by using dynamic manufacturing conditions along with static machining parameters. Experiments are conducted to evaluate MTF-CLSTM performance to show

that MTF-CLSTM significantly outperforms the existing method in terms of the prediction mean absolute percentage error (MAPE).

The rest of this paper is organized as follows. Section 2 introduces some background knowledge, and Section 3 shows some related research results. The proposed method is elaborated in Section 4. Performance evaluation and comparisons of the proposed method with related methods are demonstrated in Section 5. Finally, concluding remarks are drawn in Section 6.

## 2. Background Knowledge

This section describes some background knowledge, including the MTF, CNN, and LSTM models. Below, the models are introduced one by one in separate subsections.

### 2.1. Markov Transition Field (MTF)

The Markov transition field (MTF) is closely related to the Markov chain, as introduced below. The Markov chain can be used to model the state-to-state transitions of a system [32]. It uses the state transition diagram or state transition matrix (also called Markov transition matrix) to describe the probabilities of a state transitioning to itself or other states. For example, Figure 2 is the state transition diagram corresponding to a 4-state Markov chain.

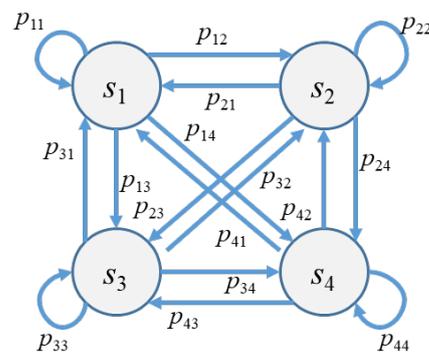


Figure 2. The 4-state Markov chain with 16 state-transition probabilities.

Below is the Markov transition matrix  $P_{44}$  corresponding to the Markov transition diagram shown in Figure 2. In the diagram and the matrix,  $s_1, s_2, s_3,$  and  $s_4$  are the four states, and  $p_{ij}$  is the probability of state  $s_i$  transitioning to state  $s_j$ , where  $\sum_i p_{ij} = 1$ , and  $1 \leq i, j \leq 4$ .

$$P_{44} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \tag{1}$$

In general, an  $m$ -state Markov chain with states  $s_1, \dots, s_m$  can be represented by an  $m \times m$  Markov transition matrix  $P_{mm}$ , where  $p_{ij}$  is the probability of state  $s_i$  transitioning to state  $s_j$ ,  $\sum_i p_{ij} = 1$ , and  $1 \leq i, j \leq m$ , as shown below.

$$P_{mm} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \tag{2}$$

Wang and Oates introduced the concept of the Markov transition field [29], as described below. Given a time series  $X = (x_1, x_2, \dots, x_n)$ , a data point  $x_t$  at time step  $t$  ( $1 \leq t \leq n$ ) is first assigned to a corresponding state  $s_j$  (or a quantile bin  $q_j$ ), where  $1 \leq j \leq m$ , and  $m$  is the number of states (or quantile bins). In this way, an  $m \times m$  Markov transition matrix  $P_{mm}$  associated with the time series  $X$  can be derived by first

calculating  $c_{ij}(1 \leq i, j \leq m)$ , which is the count of data points in state  $s_i$  transiting to state  $s_j$ . Afterwards, each entry  $p_{ij}$  of  $P_{mm}$  can be derived as  $p_{ij} = \frac{c_{ij}}{\sum_i c_{ij}}$ . It can easily check that  $\sum_i p_{ij} = 1$ . The Markov transition field in practice captures the multi-span transition probability between any two data points in  $X = (x_1, x_2, \dots, x_n)$ . It is an  $n \times n$  matrix, as given below.

$$MTF_{nm} = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ f_{21} & \cdots & f_{2n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix} = \begin{bmatrix} p_{ij|x_1 \in s_i, x_1 \in s_j} & \cdots & p_{ij|x_1 \in s_i, x_n \in s_j} \\ p_{ij|x_2 \in s_i, x_1 \in s_j} & \cdots & p_{ij|x_2 \in s_i, x_n \in s_j} \\ \vdots & \ddots & \vdots \\ p_{ij|x_n \in s_i, x_1 \in s_j} & \cdots & p_{ij|x_n \in s_i, x_n \in s_j} \end{bmatrix} \quad (3)$$

In the above equation,  $f_{kl} = p_{ij}$  ( $1 \leq k, l \leq n$ , and  $1 \leq i, j \leq m$ ) is the probability that state  $s_i$  of data point  $x_k$  at time step  $k$  transits to state  $s_j$  of data point  $x_l$  at time step  $l$ . Compared with the Markov transition matrix, the Markov transition field has extra temporal information besides state transition probabilities. It is thus more suitable for representing and extracting features of time series. For a time series of a large number  $n$  of data points, its associated Markov transition field is a large  $n \times n$  matrix, which is usually regarded as an image for the purpose of analyzing and visualizing.

### 2.2. Convolutional Neural Network (CNN)

The convolutional neural network (CNN) is a powerful and efficient artificial neural network with the characteristics of neural parameter sharing and sparsity of neural connections. As shown in Figure 3, a CNN usually takes an image as input and contains the input layer, several groups of convolutional layers and pooling layers, one or more fully connected layers, and the output layer [30].

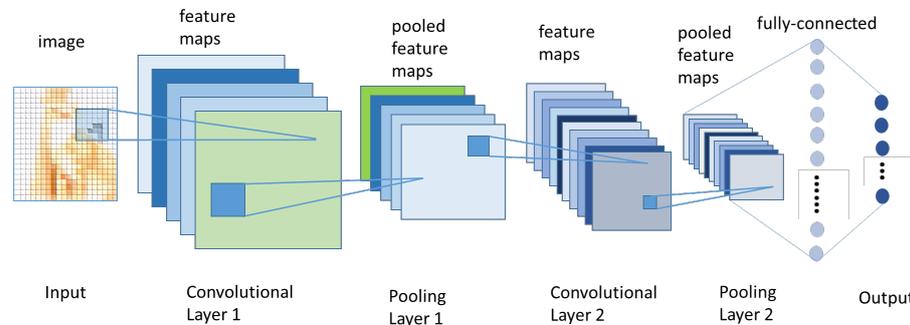


Figure 3. The network architecture of a CNN.

Filters (or kernels) are applied in convolutional layers to slide over the image to perform the convolution operation for extracting image features, which are called feature maps. One filter generates one feature map, corresponding to a channel to be fed into the following layer. Note that a non-linear activation function, such as the rectified linear unit (ReLU) function, is applied to the feature map before it is generated. Filters have different sizes and different hyperparameters, such as the stride and the padding. A filter with width  $w$ , length  $l$ , stride  $s$ , and padding  $p$  slides over the image in the left-to-right and top-to-bottom manner. When the filter moves, it jumps  $s$  pixels for every move, with  $p$  of zeros are padded on the image borders. Filters are also used in the pooling layer to slide over image maps for the purpose of subsampling the image maps (i.e., reducing the image map sizes) while maintaining critical image map features. The maximum pooling layer, which returns the maximum value in the filter region, and the average pooling layer, which returns the average of values in the filter region, are two typical pooling layers. Fully connected layers (or dense layers) come after the convolutional and the pooling layers.

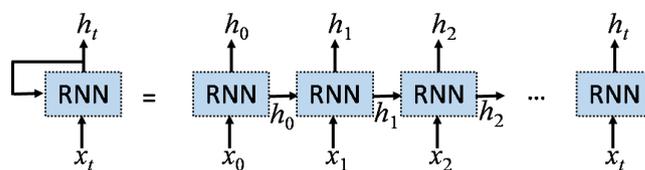
For connecting to the dense layer, the image maps generated by the last pooling layer are flattened; that is, they are transformed from the multiple-dimension shape into the one-dimension shape and concatenated as a multi-tuple vector. The dropout mechanism is usually applied in the dense layers to avoid over-fitting. The multiple-tuple vector then goes through zero, one or more dense layers, and finally the output layer. The softmax function is used in the output layer when the CNN is for the purpose of classifying the input image. However, another activation function, such as the sigmoid function, is used when the CNN is for the purpose of generating values associated with the input image.

### 2.3. Long Short-Term Memory (LSTM) Neural Network

The long short-term memory (LSTM) neural network [31] is a well-known and effective deep-learning-based model. It is a special type of recurrent neural networks (RNN) [33]. RNNs are suitable for processing time series, as they add a loop to a neuron allowing the output at the current time point to be used as the input at the next time point. Figure 4 shows the structure of the RNN and its corresponding logical structure for processing time series. In Figure 4,  $x_t$  and  $h_t$  are the input data and the output (or hidden state) at time point  $t$ , respectively. The equation associated with the RNN is shown below.

$$h_t = \sigma(W \cdot [h_{t-1} || x_t] + b), \tag{4}$$

where  $\sigma(\ )$  stands for an activation function,  $W$  stands for weights,  $||$  stands for the concatenation operation,  $b$  stands for the bias. The RNN behaves as if it could memorize previous input values and output results to generate the current output. However, it has the problems of gradient vanishing and gradient exploding when tuning neural network link weights with the traditional gradient descent error backpropagation mechanism. Therefore, the RNN is difficult to reflect the dependency of input data that are separated far apart in the time series.



**Figure 4.** The structure of the RNN (the left part to the equation sign) and its logical structure for processing time series (the right part to the equation sign).

The LSTM neural network [31] can mitigate the gradient vanishing and gradient exploding problems of the RNN by including in an LSTM unit a memory cell and three gates: the input gate, the output gate, and the forget gate. Please refer to Figure 5 for the details of the unit structure of an LSTM neural network, by which information can be added to or removed from the memory cell via the control of the gates. The weights associated with gates can be learned so that the memory cell can store the most necessary historical information to produce the most proper output. Six Equations (5)–(10) associated with the LSTM neural network are shown below. They are for the forget gate, input gate, intermediate value of memory cell, memory cell, output gate, and output (hidden state), respectively. In the equations,  $||$  stands for the concatenation operation,  $\times$  stands for the Hadamard product (element-wise product) operation,  $W$  stands for weights,  $b$  stands for the bias,  $\sigma(\ )$  stands for the sigmoid function, and  $\tanh(\ )$  stands for the hyperbolic tangent function.

$$f_t = \sigma(W_f \cdot [h_{t-1} || x_t] + b_f) \tag{5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1} || x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1} || x_t] + b_c) \tag{7}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{8}$$

$$o_t = \sigma(W_o \cdot [h_{t-1} || x_t] + b_o) \quad (9)$$

$$h_t = o_t \times \tanh(C_t) \quad (10)$$

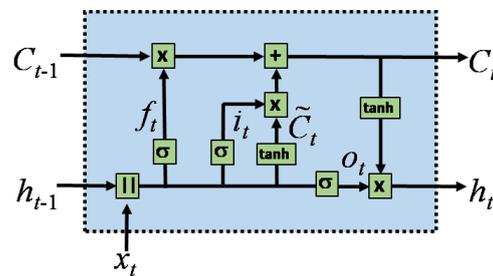


Figure 5. The unit structure of the LSTM neural network.

### 3. Related Work

This section reviews some research studies [19–28] related to WEDM product surface roughness prediction. They applied different prediction methods to different WEDM machines with varying machining parameters for machining workpieces of various materials. Below, the studies are elaborated one by one.

The study [19] used two methods, namely the linear regression (LR) and the artificial neural network (ANN) with only one hidden layer, to predict WEDM surface roughness of product of AISI 4340 steel. The WEDM machine was the Accutex GE series [34] using the wire electrode of brass with the diameter of 0.25 mm. Four machining parameters, the pulse-on time, open voltage, wire feed rate, and dielectric flushing pressure, were used as inputs of the two methods. The MAPEs of the surface roughness prediction of the two methods were 7.17% and 4.94%, respectively.

The research [20] used the 2nd-order regression method to predict the surface roughness of the grade-2 titanium workpiece machined by the Electronica Sprintcut 734 WEDM machine [35]. The electrode was the brass wire with the diameter of 0.25 mm. The Box-Behnken design for the response surface methodology (RSM) was applied for setting up experiments. Six machining parameters, the pulse-on time, pulse-off time, peak current, gap voltage, wire feed, and wire tension, were used as inputs of the method for the purpose of predicting WEDM product surface roughness. The MAPE of the prediction was 3%.

The research [21] focused on the prediction and the comparison of WEDM performance for Al7075-TiB<sub>2</sub> (Alumine 7075-Titanium di-boride) in-situ composite in terms of the surface roughness, material removal rate, and dimensional error. The machining parameters considered by the research are pulse-on time, pulse-off time, and current bed speed. They were selected based on the Taguchi L27 orthogonal array. The ANN model was used for predicting WEDM performances. The minimum and the maximum deviation between the measured and the predicted surface roughness were 1.3% and 12.51% respectively.

The research [22] estimated the surface roughness, accuracy, material removal rate and electrode wear for workpiece material of Al(5% wt)-Si<sub>3</sub>N<sub>4</sub> based on machining parameters, such as the pulse-on time, pulse-off time, current and bed speed. The Taguchi L27 orthogonal array was used to select parameters and the ANN model was used to perform the prediction. The experiments were carried out on the Concord DK7720C WEDM machine [36] using the molybdenum wire of 0.18 mm diameter as an electrode. The predicted results were shown to coincide with the measured results.

The research [23] proposed a WEDM machining quality prediction method for workpieces of Inconel 718 material in terms of the surface roughness, cutting speed, material removal rate, and sparking gap. The prediction method was based on the cascade forward neural network (CFNN) considering five machining parameters like the pulse-on time, pulse-off time, peak current, servo voltage, and flushing pressure. The Taguchi L256 orthogonal array was applied to setting machining parameter level combinations. The experiments were realized on Sodick AQ537L WEDM machine [37] using the brass

wire of diameter 0.25 mm as the wire electrode. The surface roughness prediction MAPE of the method was 2.00%.

The research [24] used the Electronica Ultracut S0 WEDM machine [35] to conduct machining experiments on workpieces of the Al 2124 SiCp (0% wt, 15% wt, 20% wt) metal matrix composite (MMC) material for performing dimensional analysis (DA) and for modeling an ANN to predict the workpiece surface roughness and the material removal rate. Machining parameters, such as the pulse-on time, pulse-off time, duty cycle, wire feed rate, wire tension, peak current, and gap voltage were taken as inputs. Furthermore, the density, thermal conductivity, thermal expansion, SiC powder weight of the workpiece material were also taken as inputs. The predicted surface roughness by the DA and the ANN were of correlation coefficients ( $R^2$ ) of 0.92345 and 0.99999, respectively.

In research [25], a WEDM machine, Agie Charmilles CUT 20P [38], was used for machining workpieces of materials of Al-Sn-SiC MMC with varying Sn and SiC weight percentages (5%, 10%, 15%, and 20% Sn wt% and SiC wt%) alloyed into aluminum. The workpiece was of 6 mm height and the brass wire of 0.25 mm diameter was used as the wire electrode. Machining parameters, such as the pulse-on time, pulse-off time, wire feed rate, along with Sn wt% and SiC wt%, were taken as inputs of an ANN for predicting the surface roughness of the machining. The Taguchi L32 orthogonal array was used for the experimental design of machining parameter combinations. The ANN model predicted the WEDM workpiece surface roughness with the correlation coefficient ( $R^2$ ) of 0.9851.

The research [26] used the support vector machine (SVM) to predict the surface roughness of WEDM workpieces. The workpiece was of material AA6063, which is an Al-Si-Mg based alloy, and is with the size of 150 mm by 100 mm by 15 mm. Four machining parameters were taken as SVM inputs; they were the pulse-on time, pulse-off time, servo voltage, and peak current. Experiments were designed according to the full factorial design with 3 levels. The mean square error (MSE) and correlation coefficient  $R^2$  of the prediction were 0.389178  $\mu\text{m}$  and 0.963426, respectively.

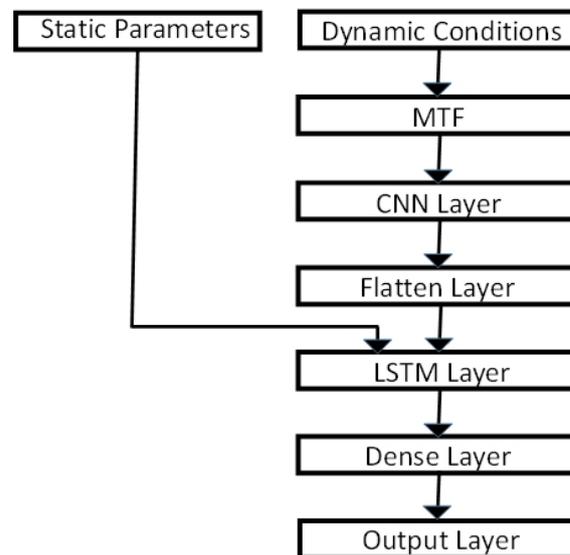
Two methods, based on the deep neural network (DNN) and the Markov chain deep neural network (MC-DNN), were proposed in [27] to predict the surface roughness of workpieces machined by the Chmer Q4025L WEDM machine [39]. The first method took static machining parameters like the pulse-on time, pulse-off time, open voltage, servo voltage, and wire tension, as inputs to perform prediction before machining. The second method took the above-mentioned static parameters along with dynamically-changing machining conditions to perform prediction after machining. The conditions were the gap voltage, servo feed rate, normal-state count, and abnormal-state count. The condition data were regarded as time series and modeled by the Markov chain to derive the Markov transition matrix as features to be fed into the DNN for predicting the workpiece surface roughness. The MAPEs of the predictions of the two methods were 4.9% and 4.68%, respectively.

The research [28] conducted experiments of machining workpieces of Al7075 aluminum alloy via Sodick SL400Q WEDM machine [37]. It proposed using four methods, the support vector regression (SVR), quadratic support vector regression (Q-SVR), extreme learning machine (ELM), and weighted extreme learning machine (W-ELM), to predict workpiece surface roughness based on machining parameters, like the pulse-on-time, open voltage, dielectric flushing pressure, and wire feed. The prediction correlation coefficient  $R^2$  of the four methods were 0.8824, 0.9613, 0.9411, and 0.9720, respectively. It was shown that the W-ELM model had the best prediction performance.

#### 4. The Proposed Method

The proposed method, called MTF-CLSTM, integrates the MTF model and the CLSTM neural network for WEDM product quality prediction right after manufacturing. The framework of the proposed MTF-CLSTM method is shown in Figure 6. Note that the combination of the CNN and the LSTM neural network is called the CLSTM neural network. This is why the proposed method is called the MTF-CLSTM method. MTF-CLSTM uses the MTF model to represent dynamic WEDM manufacturing conditions as images. It then uses the

CNN network to extract features of the images. The LSTM network then takes the features along with static machining parameters as the input data for predicting WEDM workpiece surface roughness. Note that we use the LSTM network rather than other models, as it is useful for identifying the relationship between input data points that may be far apart.

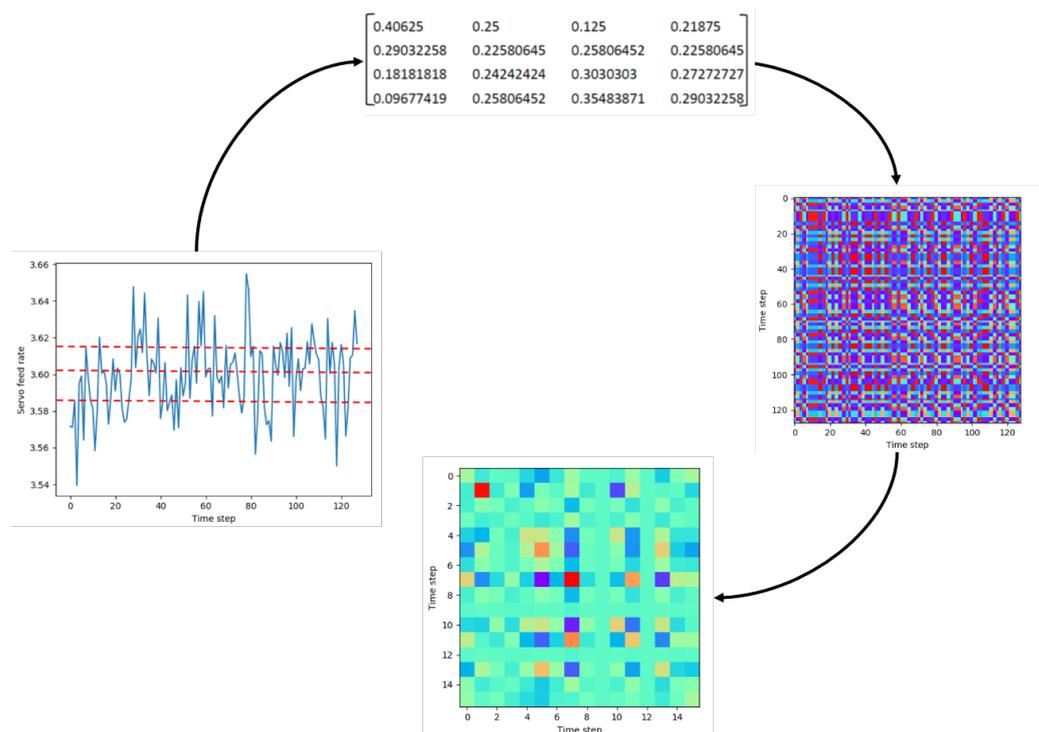


**Figure 6.** The framework of the proposed MTF-CLSTM method.

The MTF-CLSTM method takes five static machining parameters and four dynamically-changing machining conditions as inputs to perform prediction right after machining. The static machining parameters are the pulse-on time, pulse-off time, open voltage, servo voltage, and wire tension, whereas the dynamically-changing machining conditions are the gap voltage, servo feed rate, normal-state count, and abnormal-state count. The proposed MTF-CLSTM is elaborated below.

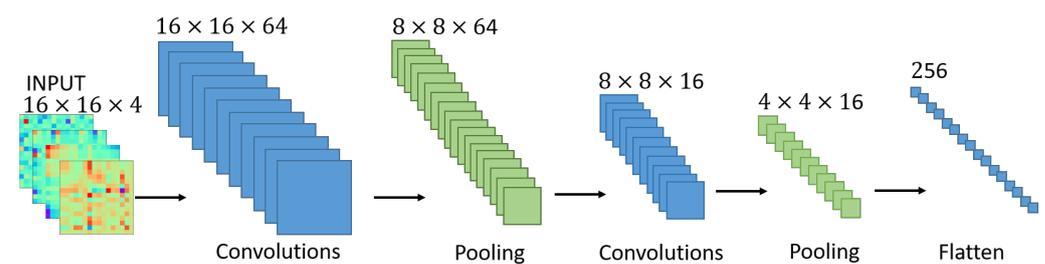
First, dynamic WEDM manufacturing condition data are first fed into the MTF model to be represented as images. Four dynamically-changing machining conditions are used by the proposed method. Hence there are four sets of data, each of which corresponds to a time series. Each time series is transformed by the MFT model to be an image. Figure 7 shows the process of the MTF transform. As shown in Figure 7, data points in a time series are classified into quantile bins (or states). The Markov transition matrix of state transition probabilities associated with the time series is then derived. The matrix is afterward used to derive the Markov transition field that is equivalently an image, called an MTF image. To reduce the size, the image further goes through a blurring process, which is similar to downsampling of the average pooling, to obtain a blurred image, called a blurred MTF image for the sake of analysis efficiency. Note that the Markov transition field contains extra temporal features besides the state transition probability features.

After the four sets of dynamically-changing machining conditions are transformed as four MTF images, the four images are fed into the CNN as a four-channel image to extract more-detailed features. The CNN used in the proposed method is shown in Figure 8. Specifically, the CNN takes images of the  $16 \times 16 \times 4$  shape as inputs. Its first convolutional layer has 64 filters of size  $3 \times 3$  with the stride of 1 and ‘same’ padding (i.e., to padding proper number of zeros to keep feature images and original images to have the same size). The first pooling layer is an average pooling layer using filters of the size  $2 \times 2$  with the stride of 2. The second convolutional layer has 16 filters of size  $3 \times 3$  with the stride of 1 and ‘same’ padding. The second pooling layer is also an average pooling layer using filters of the size  $2 \times 2$  with the stride of 2 and ‘same’ padding. Note that the LeakyReLU function is used in both the first and the second convolutional layers. After the second pooling layer, there are image maps of the  $4 \times 4 \times 16$  shape, which are flattened as a 256-tuple vector.

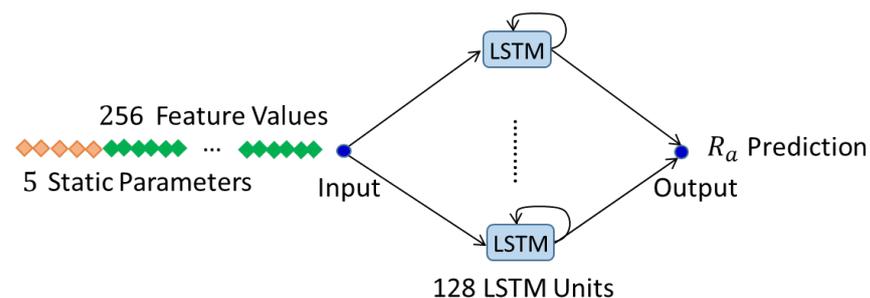


**Figure 7.** The process of transforming a time series into a Markov transition field: (left) time series and different states, (top) Markov transition matrix, (right) Markov transition field, and (bottom) blurred Markov transition field.

The flattened 256-tuple vector and the 5 static machining parameters are combined as a 261-tuple vector to be fed into the LSTM neural network for predicting the WEDM workpiece surface roughness. As shown in Figure 9, the LSTM neural network adopted by the MTF-CLSTM method contains 128 LSTM units and takes the hyperbolic tangent function as the activation function. The LSTM neural network outputs the surface roughness prediction after the 261-tuple vector is entirely fed into it.



**Figure 8.** The CNN architecture used in the proposed MTF-CLSTM method.



**Figure 9.** The LSTM architecture used in the proposed MTF-CLSTM method.

## 5. Performance Evaluation and Comparison

Experiments are conducted to evaluate the performance of the proposed MTF-CLSTM method for comparisons with related methods [19–28]. The experiments are performed on the Chmer Q4025L WEDM machine [39] for machining SKD61 steel with the length of 10 mm, the width of 10 mm, and the height of 30 mm. The brass wire with the diameter of 0.25 mm is used as the wire electrode. Totally 195 data (or files) are gathered, which are divided into training dataset of 185 data and test dataset of 10 data. Five static manufacturing parameters and four dynamic manufacturing conditions are recorded in each of the data. Each datum (or file) can be regarded as a time series of the length of 128.

The 128 data points of a data file are fed into the MFT model to be represented as a  $128 \times 128$  image. The image is downsized to be a  $16 \times 16$  blurred image, called a blurred MFT. Note that the number of states (or quantile bins) of the MFT model is taken as 3, 4, or 5. By the MFT model, 4 images of the  $16 \times 16$  size are generated, each corresponds to a dynamic manufacturing condition. The 4 images are then fed into the CNN for feature extraction with each image as a channel. The CNN layers include two convolutional layers and two pooling layers. The first convolutional layer has 64 filters, all with the size  $3 \times 3$  and the stride 1, followed by an average pooling layer with the pool size of  $2 \times 2$ . The second convolutional layer has 16 filters, all with the size  $3 \times 3$  and the stride 1, followed by an average pooling layer with the pool size  $2 \times 2$ . The activation function for each convolutional layer neuron is LeakyReLU. After the CNN flatten layer, 256 features are extracted. The 256 features, along with the 5 static manufacturing parameters, are then fed into the LSTM neural network with 128 LSTM units for identifying time-dependency relationship between features to predict the product surface roughness Ra of the WEDM workpiece. When training the CLSTM model, the batch size is taken as 32 with 150 epochs. Furthermore, the 10-fold cross validation and the early stopping mechanisms are used to avoid over-fitting the model.

The experimental results show that the mean absolute percentage errors (MAPE) of MTF-CLSTM method using 3-, 4-, and 5-state MTF are 3.11%, 2.94%, and 3.24%, respectively. Figure 10 shows the prediction versus the fact (ground truth) of the WEDM product Ra in the unit of  $\mu\text{m}$ .

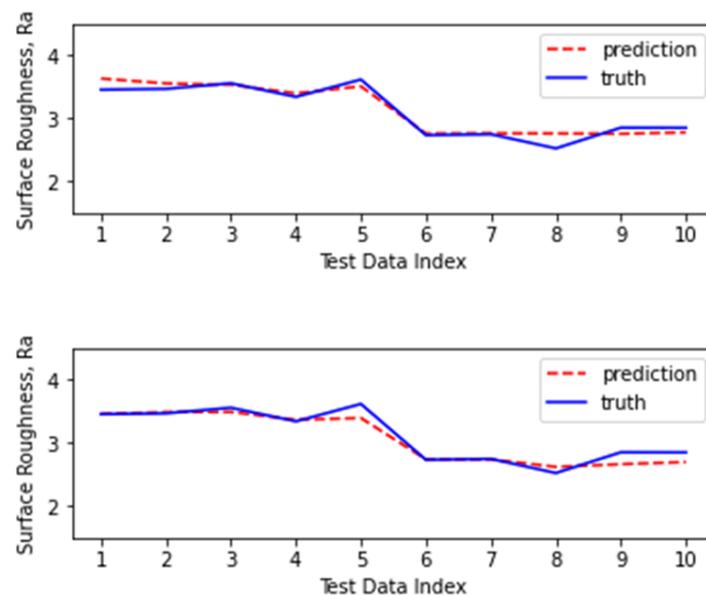
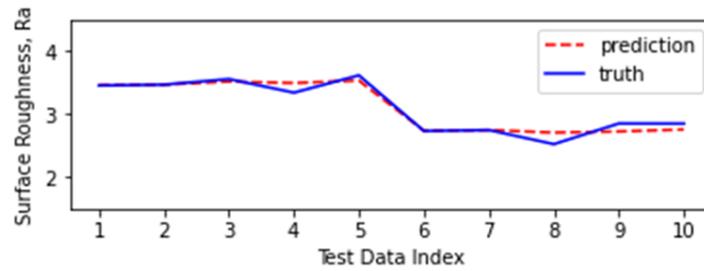


Figure 10. Cont.



**Figure 10.** The prediction and the fact of the surface roughness Ra in the unit of m for 10 testing data for the proposed MTF-CLSTM method with 3-, 4-, and 5-state MTF.

Table 1 shows the comparisons of the proposed method with related methods [19–28] predicting WEDM workpiece surface roughness. Among the compared methods, the MC-DNN method proposed in [27] is the only method that is like MTF-CLSTM to predict WEDM workpiece surface roughness by using static machining parameters and dynamic manufacturing conditions. By Table 1, it can be observed that the proposed MTF-CLSTM method significantly outperforms the MC-DNN method for all the cases of the 3-, 4-, and 5-state MTFs in terms of surface roughness prediction error MAPEs.

**Table 1.** Comparisons of the proposed method with related research predicting WEDM surface roughness.

Research	Year	Material	Machine	Workpiece Size (mm)	Parameters	Data Size	Method	Performance
Esme et al. [19]	2009	AISI 3440 Steel	Accutex GE	150 × 150 × 10	TON, OV, WF, FP	28	1. LR 2. ANN	1. MAPE: 7.17% 2. MAPE: 4.94%
Kumar et al. [20]	2012	Titanium (grade-2)	Electronica Sprintcut 734	148 × 148 × 26	TON, TOFF, PC, GV, WF, WT	54	2nd Regression	MAPE: 3%
Surya et al. [21]	2017	Al7075-TiB <sub>2</sub> MMC	–	–	TON, TOFF, PC, BS	27	ANN	Min Dev: 1.3% Max Dev: 12.51%
Gurupavan et al. [22]	2017	Al-5 wt% Si <sub>3</sub> N <sub>4</sub> MMC	Concord DK7720C	–	TON, TOFF, PC, BS	27	ANN	–
Yusoff et al. [23]	2018	Inconel 718	Sodick AQ537L	48 × 25 × 12.5	TON, TOFF, PC, SV, FP	22	CFNN	PE: 2.00%
Phate et al. [24]	2019	Al 2124 SiCp MMC	Electrinica Ultracut S0	80 × 55 × 20	TON, TOFF, WF, PC	27	1. DA 2. ANN	1. R <sup>2</sup> : 0.92345 2. R <sup>2</sup> : 0.99999
Thankachan et al. [25]	2019	Al-Sn-SiC MMC	Agie Charmilles CUT 20P	Height: 6	TON, TOFF, WF, Sn wt%, SiC wt%	32	ANN	R <sup>2</sup> : 0.9851
Singh et al. [26]	2019	AA6063	–	150 × 100 × 15	TON, TOFF, PC, SV	81	SVM	MSE: 0.389178 μm R <sup>2</sup> : 0.963426
Fan et al. [27]	2019	SKD61 Steel	Chmer Q4025L	10 × 10 × 30	Static: TON, TOFF, OV, SV, WT Dynamic: GV, SF, NS, AS	110	1. DNN(static) 2. MC-DNN (static + dynamic)	1. MAPE: 4.9% 2. MAPE: 4.68%
Ulas et al. [28]	2020	Al7075	Sodick SL400Q	Circle (φ = 10)	OV, TON, FP, WF	81	1. SVR 2. Q-SVR 3. ELM 4. W-ELM	1. R <sup>2</sup> : 0.8824 2. R <sup>2</sup> : 0.9613 3. R <sup>2</sup> : 0.9411 4. R <sup>2</sup> : 0.9720
This paper	2021	SKD61 Steel	Chmer Q4025L	10 × 10 × 30	Static: TON, TOFF, OV, SV, WT Dynamic: GV, SF, NS, AS	110	MTF-CLSTM (static + dynamic): 1. 3-state 2. 4-state 3. 5-state	1. MAPE: 3.11% 2. MAPE: 2.94% 3. MAPE: 3.24%

TON: pulse-on time; TOFF: pulse-off time; OV: open voltage; FP: dielectric flush pressure; PC: peak current; SV: servo voltage; WT: wire tension; WF: wire feed rate; GV: gap voltage; SF: servo feed rate; BS: bed speed; NS: normal-state count; AS: abnormal-state count.

## 6. Conclusions

This paper proposes a method called MTF-CLSTM, to integrate the MTF model, the CNN, and the LSTM neural network for WEDM product quality prediction right after machining. MTF-CLSTM first uses the MTF model to transform the gathered data into images to extract temporal information and state transition probability information. It further uses the CNN to extract more detailed spacial features from images. Finally, the LSTM neural network is used to capture temporal relationship that may be separated far apart in data. Experiments are conducted to evaluate the performance of the proposed method. The prediction MAPEs of the proposed method using 3-, 4-, and 5-state MTF are 3.11%, 2.94%, and 3.24%, respectively. It can be observed that MTF-CLSTM outperforms DNN and MC-DNN, which are two methods using the same experimental settings as MTF-CLSTM. Besides performance, the proposed method is also compared with related research [19–28] in many other aspects, such as the WEDM machine used, workpiece material, workpiece size, parameters used for the prediction, and so on.

In the future, the authors plan to apply the proposed MTF-CLSTM method to predict different product quality like the dimension error and the material removal rate. The authors also plan to apply hyperparameter optimization techniques [40], such as Bayesian optimization and its variants, multi-bandit mechanisms, and population based training (PBT) approaches, for facilitating hyperparameter tuning and for improving performance. The hyperparameters for tuning include the filter size, the number of filters, the number of layers, the number of neurons per layer, the dropout rate, various activation functions, and so on.

**Author Contributions:** J.-R.J. and C.-T.Y. altogether designed the proposed method. C.-T.Y. implemented the method and did performance evaluation for it. J.-R.J. wrote the paper reporting the background, related work, the proposed method, and the performance evaluation and comparisons with other related methods. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology (MOST), Taiwan, under the grant number 109-2622-E-008-028-.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Jiang, J.R. An improved cyber-physical systems architecture for Industry 4.0 smart factories. *Adv. Mech. Eng.* **2018**, *10*, 1687814018784192. [[CrossRef](#)]
2. Shivade, A.; Kubade, P.R.; Shinde, G. Multi-parametric optimization of WEDM process using desirability function analysis. *Int. Adv. Res. J. Sci. Eng. Technol.* **2017**, *4*. [[CrossRef](#)]
3. Singh, B.; Misra, J. A critical review of wire electric discharge machining. In *DAAAM International Scientific Book*; DAAAM International: Vienna, Austria, 2016.
4. Sonawane, S.A.; Kulkarni, M. Optimization of machining parameters of WEDM for Nimonic-75 alloy using principal component analysis integrated with Taguchi method. *J. King Saud Univ. Eng. Sci.* **2018**, *30*, 250–258. [[CrossRef](#)]
5. Vijayabhaskar, S.; Rajmohan, T. Experimental investigation and optimization of machining parameters in WEDM of nano-SiC particles reinforced magnesium matrix composites. *Silicon* **2019**, *11*, 1701–1716. [[CrossRef](#)]
6. Kavimani, V.; Prakash, K.S.; Thankachan, T. Multi-objective optimization in WEDM process of graphene–SiC-magnesium composite through hybrid techniques. *Measurement* **2019**, *145*, 335–349. [[CrossRef](#)]
7. Thangaraj, M.; Annamalai, R.; Moiduddin, K.; Alkindi, M.; Ramalingam, S.; Alghamdi, O. Enhancing the surface quality of micro titanium alloy specimen in WEDM process by adopting TGRA-based optimization. *Materials* **2020**, *13*, 1440. [[CrossRef](#)]
8. Kulkarni, V.N.; Gaitonde, V.; Karnik, S.; Manjaiah, M.; Davim, J.P. Machinability analysis and optimization in wire EDM of medical grade NiTiNOL memory alloy. *Materials* **2020**, *13*, 2184. [[CrossRef](#)]
9. Ezeddini, S.; Boujelbene, M.; Bayraktar, E.; Ben Salem, S. Optimization of the Surface Roughness Parameters of Ti–Al Intermetallic Based Composite Machined by Wire Electrical Discharge Machining. *Coatings* **2020**, *10*, 900. [[CrossRef](#)]
10. Sibalija, T.V.; Kumar, S.; Patel, G.M. A soft computing-based study on WEDM optimization in processing Inconel 625. *Neural Comput. Appl.* **2021**, 1–22. [[CrossRef](#)]
11. Modrak, V.; Pandian, R.S.; Kumar, S.S. Parametric Study of Wire-EDM Process in Al-Mg-MoS<sub>2</sub> Composite Using NSGA-II and MOPSO Algorithms. *Processes* **2021**, *9*, 469. [[CrossRef](#)]

12. Phate, M.R.; Toney, S.B.; Phate, V.R. Multi-parametric Optimization of WEDM Using Artificial Neural Network (ANN)-Based PCA for Al/SiCp MMC. *J. Inst. Eng.* **2021**, *102*, 169–181.
13. Bose, S.; Nandi, T. Parametric optimization of WEDM on hybrid titanium matrix composite using response surface methodology. *Multiscale Multidiscipl. Model. Exper. Des.* **2021**, 1–8. [[CrossRef](#)]
14. Pu, Y.; Zhao, Y.; Meng, J.; Zhao, G.; Zhang, H.; Liu, Q. Process Parameters Optimization Using Taguchi-Based Grey Relational Analysis in Laser-Assisted Machining of Si<sub>3</sub>N<sub>4</sub>. *Materials* **2021**, *14*, 529. [[CrossRef](#)]
15. Kachhap, S.; Singh, A. Prediction of controllable process variables for various workpiece materials in CNC-WEDM. In *Innovation in Materials Science and Engineering*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 131–142.
16. Naresh, C.; Bose, P.; Rao, C. Artificial neural networks and adaptive neuro-fuzzy models for predicting WEDM machining responses of Nitinol alloy: Comparative study. *SN Appl. Sci.* **2020**, *2*, 1–23. [[CrossRef](#)]
17. Chalisgaonkar, R.; Kumar, J.; Pant, P. Prediction of machining characteristics of finish cut WEDM process for pure titanium using feed forward back propagation neural network. *Mater. Today Proc.* **2020**, *25*, 592–601. [[CrossRef](#)]
18. Lalwani, V.; Sharma, P.; Pruncu, C.I.; Unune, D.R. Response Surface Methodology and Artificial Neural Network-Based Models for Predicting Performance of Wire Electrical Discharge Machining of Inconel 718 Alloy. *J. Manuf. Mater. Process.* **2020**, *4*, 44. [[CrossRef](#)]
19. Esme, U.; Sagbas, A.; Kahraman, F. Prediction of surface roughness in wire electrical discharge machining using design of experiments and neural networks. *Iranian J. Sci. Technol. Trans. B Eng.* **2009**, *33*, 231–240.
20. Kumar, A.; Kumar, V.; Kumar, J. Prediction of surface roughness in wire electric discharge machining (WEDM) process based on response surface methodology. *Int. J. Eng. Technol.* **2012**, *2*, 708–719.
21. Surya, V.R.; Kumar, K.V.; Keshavamurthy, R.; Ugrasen, G.; Ravindra, H. Prediction of machining characteristics using artificial neural network in wire EDM of Al7075 based in-situ composite. *Mater. Today Proc.* **2017**, *4*, 203–212. [[CrossRef](#)]
22. Gurupavan, H.; Devegowda, T.; Ravindra, H.; Ugrasen, G. Estimation of machining performances in WEDM of aluminium based metal matrix composite material using ANN. *Mater. Today Proc.* **2017**, *4*, 10035–10038. [[CrossRef](#)]
23. Yusoff, Y.; Zain, A.M.; Sharif, S.; Sallehuddin, R.; Ngadiman, M.S. Potential ANN prediction model for multiperformances WEDM on Inconel 718. *Neural Comput. Appl.* **2018**, *30*, 2113–2127. [[CrossRef](#)]
24. Phate, M.R.; Toney, S.B. Modeling and prediction of WEDM performance parameters for Al/SiCp MMC using dimensional analysis and artificial neural network. *Eng. Sci. Technol. Int. J.* **2019**, *22*, 468–476. [[CrossRef](#)]
25. Thankachan, T.; Prakash, K.S.; Malini, R.; Ramu, S.; Sundararaj, P.; Rajandran, S.; Rammasamy, D.; Jothi, S. Prediction of surface roughness and material removal rate in wire electrical discharge machining on aluminum based alloys/composites using Taguchi coupled Grey Relational Analysis and Artificial Neural Networks. *Appl. Surf. Sci.* **2019**, *472*, 22–35. [[CrossRef](#)]
26. Singh, T.; Kumar, P.; Misra, J.P. Surface roughness prediction modelling for wedm of aa6063 using support vector machine technique. *Mater. Sci. Forum.* **2019**, *969*, 607–612. [[CrossRef](#)]
27. Fan, C.L.; Jiang, J.R. Surface roughness prediction based on Markov chain and deep neural network for wire electrical discharge machining. In Proceedings of the 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 3–6 October 2019; pp. 191–194.
28. Ulas, M.; Aydur, O.; Gurgenc, T.; Ozel, C. Surface roughness prediction of machined aluminum alloy with wire electrical discharge machining by different machine learning algorithms. *J. Mater. Res. Technol.* **2020**, *9*, 12512–12524. [[CrossRef](#)]
29. Wang, Z.; Oates, T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–26 January 2015; Volume 1.
30. Albelwi, S.; Mahmood, A. A framework for designing the architectures of deep convolutional neural networks. *Entropy* **2017**, *19*, 242. [[CrossRef](#)]
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
32. Behrends, E. *Introduction to Markov Chains*; Springer: Berlin/Heidelberg, Germany, 2000; Volume 228.
33. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* **2015**, arXiv:1506.00019.
34. Accutex—GE Series. Available online: <https://www.accutex.com.tw/products.htm> (accessed on 5 April 2021).
35. Electronica—Sprintcut 734. Available online: <https://electronicagroup.com/cnc-wirecut-edm/>. (accessed on 5 April 2021).
36. Concord Wire EDM. Available online: <https://concordunited.com/> (accessed on 5 April 2021).
37. Sodick WEDM. Available online: <https://www.sodick.co.jp/> (accessed on 5 April 2021).
38. Agie Charmilles WEDM. Available online: <https://www.gfms.com/com/en.html> (accessed on 5 April 2021).
39. Chmer WEDM. Available online: <http://www.chmer.com/tw/> (accessed on 5 April 2021).
40. Yu, T.; Zhu, H. Hyper-parameter optimization: A review of algorithms and applications. *arXiv* **2020**, arXiv:2003.05689.