micromachines

MDPI

Article

Applying Acoustic Signals to Monitor Hybrid Electrical Discharge-Turning with Artificial Neural Networks

Mehdi Soleymani [1] and Mohammadjafar Hadad [1,2,*]

[1] School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran P.O. Box 14155-6619, Iran; soleimani.mehdi@ut.ac.ir
[2] Department of Mechanical Engineering, College of Engineering and Technology, University of Doha for Science and Technology, Doha P.O. Box 24449, Qatar
* Correspondence: m.j.hadad@ut.ac.ir or mohammadjafar.hadad@udst.edu.qa

Abstract: Artificial intelligence (AI) models have demonstrated their capabilities across various fields by performing tasks that are currently handled by humans. However, the training of these models faces several limitations, such as the need for sufficient data. This study proposes the use of acoustic signals as training data as this method offers a simpler way to obtain a large dataset compared to traditional approaches. Acoustic signals contain valuable information about the process behavior. We investigated the ability of extracting useful features from acoustic data expecting to predict labels separately by a multilabel classifier rather than as a multiclass classifier. This study focuses on electrical discharge turning (EDT) as a hybrid process of electrical discharge machining (EDM) and turning, an intricate process with multiple influencing parameters. The sounds generated during EDT were recorded and used as training data. The sounds underwent preprocessing to examine the effects of the parameters used for feature extraction prior to feeding the data into the ANN model. The parameters investigated included sample rate, length of the FFT window, hop length, and the number of mel-frequency cepstral coefficients (MFCC). The study aimed to determine the optimal preprocessing parameters considering the highest precision, recall, and F1 scores. The results revealed that instead of using the default set values in the python packages, it is necessary to investigate the preprocessing parameters to find the optimal values for the maximum classification performance. The promising results of the multi-label classification model depicted that it is possible to detect various aspects of a process simultaneously receiving single data, which is very beneficial in monitoring. The results also indicated that the highest prediction scores could be achieved by setting the sample rate, length of the FFT window, hop length, and number of MFCC to 4500 Hz, 1024, 256, and 80, respectively.

Keywords: hybrid processes; surface roughness; neural networks; multi-label classification; electrical discharge turning; deep learning

check for updates

Academic Editors: Haiyang Fan, Shuo Yin, Jincheng Wang, Sasan Dadbakhsh and Changjun Han

Received: 27 November 2024
Revised: 23 December 2024
Accepted: 27 December 2024
Published: 27 February 2025

Citation: Soleymani, M.; Hadad, M. Applying Acoustic Signals to Monitor Hybrid Electrical Discharge-Turning with Artificial Neural Networks. *Micromachines* **2025**, *16*, 274. https://doi.org/10.3390/mi16030274

Copyright: © 2025 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

AI models have played a significant role in enhancing the performance of various processes, including those in manufacturing [1]. AI models can help reduce manufacturing costs while enhancing process stability and product quality [2]. Support vector machines, artificial neural networks, adaptive neuro-fuzzy, and regression models can classify weld defects or predict the strength of welded joints by training with radiographic images, surface images, or other welding features [3–5]. Artificial Neural Networks have also demonstrated a strong performance in predicting beam quality [6], failure load, over-weld,

normal weld, and under-weld conditions [7], as well as the static and fatigue behavior of joints [8]. To address complex prediction problems with mathematical limitations, AI can be an effective solution [9]. AI models have been widely employed to assess mechanical properties such as hardness [10–13], elastic limit [10], yield strength [10], and residual stress [13–15]. Although most of the applied AI models utilize numerical, categorical, or image-based data for their input training, some researchers have opted for acoustic data and applied their trained models across various industries. For example, breathing sounds were used to train an AI model to predict COVID-19 [16]. Additionally, acoustic signals generated during the movement of a car wheel on the road were used to classify the wetness of road surfaces [17]. AI models can make it easier to improve vehicle safety by classifying road surface roughness [18], monitor roads by classifying asphalt damages [19], and monitor pavement quality [20], all by training AI models with acoustic signals.

Acoustic data are also valuable for training AI models used in manufacturing processes. For example, welding defects can be detected using cross-validation and support vector machines in Gas Tungsten Arc Welding (GTAW) [21]. In another study, the sound of arcs generated during MIG/MAG was utilized to train an Artificial Neural Network (ANN) model to monitor the quality and stability of the process [22]. Also, in order to automate the GTAW process and control welding penetration, the training features were extracted from acoustic signals of GTAW where an ANN model was trained as the predictor model [23]. Acoustic signals could be used for training monitoring models in other manufacturing processes. An AI model was trained to monitor the tool wear in milling operations [24] and an accurate ANN model was trained to predict the degree of tool wear in milling processes [25]. The researchers mentioned that they were able to develop a tool monitoring model with a simple structure when they used acoustic signals as the training data [26]. The depth of cut, feed rate, and rotational speed were predicted by a deep neural network model trained by the acoustic emission signals in the turning process [2]. When turning Inconel 718, the resulting surface roughness was predicted by training a convolutional neural network using signals acquired from acoustic emission [27]. Multi-Layer Perceptrons, CNNs, Long Short-Term Memory (LSTM) networks, and transformer models were trained using sound and force data generated during the machining process to classify surface roughness [28]. Additionally, feeding a convolutional neural network with sounds recorded during the grinding process enabled the intelligent model to predict conditions of the grinding wheel [29]. Tool wear was evaluated by examining the tool flank during the turning process using conventional methods, including Logistic Regression and Neural Networks. By utilizing the sounds generated during the turning process, the researchers were able to classify the condition of tool wear [30]. A similar study was conducted to monitor tool wear values by preprocessing the machining audio signals by Fast Fourier Transform and training them with a convolutional neural network model [31]. Figure 1 shows the main components of the electrical discharge turning (EDT) process, the rotating workpiece and the tool electrode. EDT is a form of electrical discharge machining (EDM) which benefits from both EMD and turning processes in machining the difficult-to-machine bars as a hybrid process [32]. Pulse-on time and peak current effect the surface roughness when machining the Ti-6Al-4V by EDT [33,34]. Among the discharge current, magnetic field, pulse on time, and rotational speed, the magnetic field had the most significant effect on the EDT results [35]. The evaluation of the effects of workpiece rotational speed, electrode shape, and jump down time on circularity deviation and surface roughness revealed that workpiece rotational speed and electrode shape had the most significant impacts on surface roughness in the EDT process. In fact, increasing the rotational speed leads to a better flushing of debris, resulting in a smoother surface [36]. The effect of rotational speed, pulse on-time, and discharge current on MRR, surface roughness, and tool wear was studied

on AISI L2 steel in the EDT process [37]. Increasing the rotational speed resulted in better flushing and an increased MRR. Additionally, higher rotational speeds improved surface roughness by reducing the recast layer and enhancing flushing [38]. When machining AISI D2 steel by incorporating a magnetic field into the EDT, the magnetic flux and rotational speed were reported as the most important parameters that affect surface integrity [39]. The optimization of accuracy and precision in the (EDT) process using Taguchi robust design revealed that the most important parameters affecting MRR are pulse-on time, servo control, intensity, and rotational speed [40].

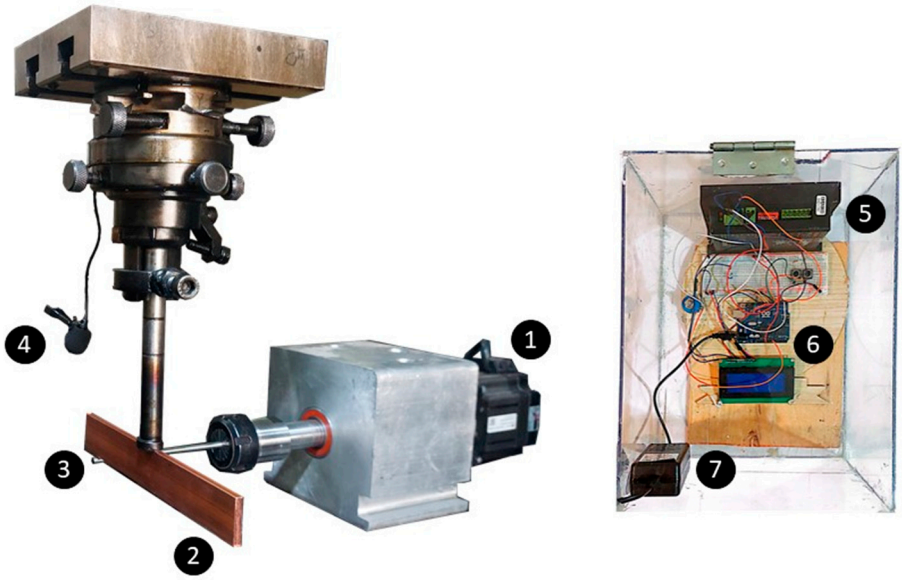


Figure 1. EDT setup, the driver and controller, and the recording device applied for machining and signal recording.

In training a conventional classification model, the model should be able to detect the only output label, which is related to the input data. But, in training a multilabel classification model, the trained model should learn how to predict the multiple output classes for a single input. For example, when predicting the gene categories in biology, the gene can be labeled as its metabolism, protein synthesis, and transcription simultaneously. If the $\mathcal{L} = \{ \lambda_1, \lambda_2, \lambda_3, \ldots \lambda_m \}$ is considered the class labels and X as an instance space in a multilabel model, any instance like $x \in X$ can be related to a subspace of labels $L \in 2^{\mathcal{L}}$, which are the subset of relevant labels, and $\mathcal{L} \backslash L$ is its complement including the irrelevant labels. After training, the multilabel classifier should be able to classify $h: X \rightarrow 2^{\mathcal{L}}$ when T is a finite set of recorded data as training data which is in the form of tuples $(x, L_x) \in X \times 2^{\mathcal{L}}$. The prediction is being performed according to the specified loss function and minimizing loss function [41].

Although there is a considerable amount of research on evaluating the ability of AI models to enhance the prediction and monitoring of manufacturing processes, researchers often struggle to provide sufficient data for training such models, which is crucial for developing an accurate and reliable model. Since preparing large datasets and numerous images as typical training data is very time-consuming and costly, we decided to extract the training data from the sounds generated during the processes. The approach of using signals like the acoustic emission signals was used by the previous researchers to detect the depth of cut, feed rate, and spindle speed in the turning process [2] or to predict the surface roughness [27], but the acoustic emission method is globally a high-cost method and needs to multiple especial sensors. So, we proposed using a microphone to record the audio-based signals as the training data which is a simple and low-cost method. Additionally, the

sounds produced by the devices provide significant information that can be easily recorded with a microphone [42]. As some researchers [40] have pointed out, the hybrid EDT process presents various parameters that need optimization; thus, we selected this process as the focus of our evaluation. The input process parameters included workpiece rotational speed, pulse on-time, and discharge current, which many researchers have identified as significant parameters affecting EDT outputs. We could have trained different AI models to predict only one factor, such as the surface roughness, using the input acoustic signals with typical multiclass classification models. In such a case, multiple models to predict each special output would have to be trained. Instead, we aimed to improve the training and predicting tasks by developing a multilabel classification model, which is rarely seen in similar manufacturing applications. This way, we could train one model which is able to predict multiple parameters simultaneously. This AI multilabel classification model is expected not only to predict the surface roughness based on the given acoustic signals, but also to detect the values of workpiece rotational speed, pulse on-time, and discharge current corresponding to the recorded acoustic signals. While predicting the surface roughness could be an alternative method to evaluate the machined surface condition, predicting the value of the mentioned input factors could help to monitor the EDT condition by a new online method using a simple microphone. The model, which could have been trained on fewer prediction labels in a standard classification model, should learn to differentiate among numerous labels in a multilabel classification setting when given a recorded acoustic signal.

2. Material and Methods

2.1. Experimental Procedure

To conduct the EDT process as a hybrid process of Turning and EDM, a rotating mechanism was mounted on an EDM machine, as shown in Figure 1. The stepper motor driver (depicted as 5) and the connected microcontroller (depicted as 6) precisely supply the required power to the stepper motor (depicted as 1) and control the set value of the rotational speed. The number 7 shows the adapter for the microcontroller. A high-speed steel bar (depicted as 3) with a 5 mm diameter was used as the workpiece. Given that natural erosion occurs either in the workpiece or the tool electrode during the EDT process, the copper strip used as the tool electrode (depicted as 2) and the workpiece were long enough to be repositioned after each test. This ensured that all tests experienced similar conditions under constant parameters. The microphone (depicted as 4), used to record sounds during the process, was a COMICA SIG.LAV V05 and was positioned near the machining area, as shown in Figure 1. Its specifications are demonstrated in Table 1.

Table 1. Microphone specifications.

Sensitivity	Gain Range	Frequency	S/N	Max SPL	Working Temperature	Storage Temperature
−35 dB ± 2 dB	Maximum +10 dB	60 Hz–18 kHz	>65 dB	100 dB	0~+50 °C	−20~+60 °C

In order to record the sound from the machining area, two different values were considered for each input parameter, specifically the workpiece rotational speed (10–20 rpm), pulse on-time (100–200 µs), and discharge current (1.5–2.5 A). All possible combinations of these parameters were evaluated, leading to eight different tests conducted in the EDT process, as shown in Table 2. During each machining test, the generated sounds were recorded, and the recorded files were split into shorter segments to prepare the data for preprocessing and feeding into the AI model. In total, out of the 3289 split acoustic files, 2631 files

were designated as training and validation data for model training, while 658 files were reserved as test data, which were not included in the training set for the final evaluation of the trained model. Additionally, the surface roughness resulting from each experiment was measured using a HOMMELWERKE Turbo Roughness Tester (V3.34), and the value recorded for each test was used as the surface roughness for each trial, as depicted in Table 2. Additionally, Figure 2 shows the surface of the workpiece machined by the EDT process which underwent surface roughness testing.

Table 2. Tests and the experimental details.

Test No	Rotational Speed (rpm)	Ton (µs)	Amperage (A)	Surface Roughness (µm)
1	10	100	2.25	2.26
2	10	100	1.5	2.22
3	10	200	2.25	2.35
4	10	200	1.5	2.29
5	20	100	2.25	1.86
6	20	100	1.5	1.86
7	20	200	2.25	2.18
8	20	200	1.5	1.97

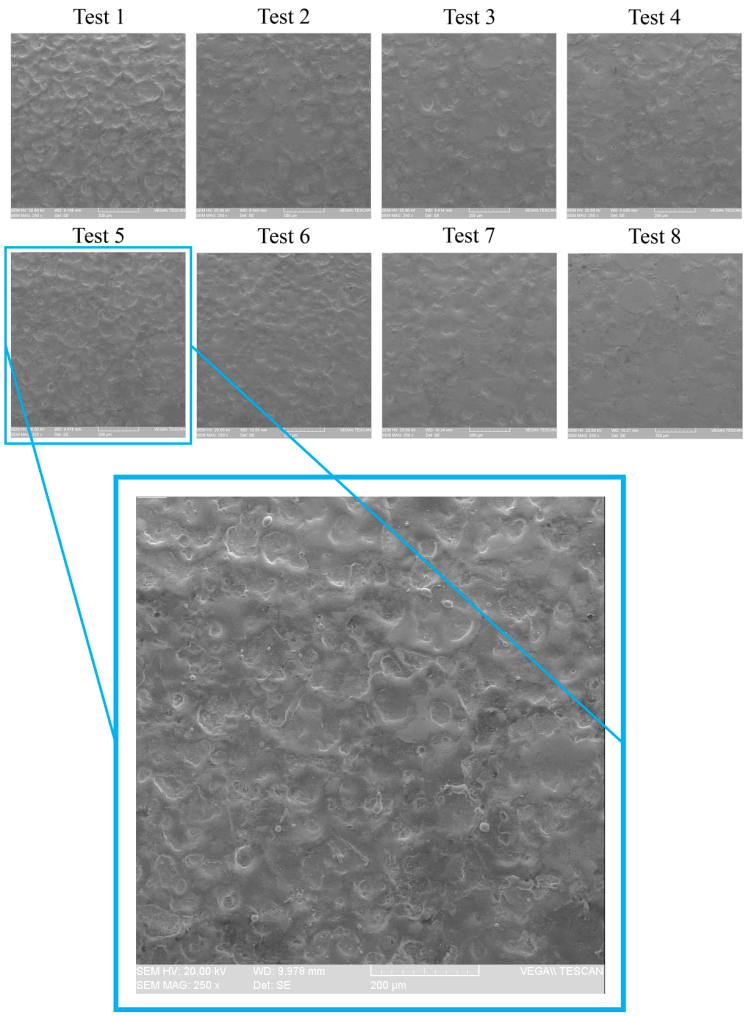


Figure 2. Scanning Electron Microscope images from the machined surface.

2.2. Input Preprocessing

To make analog acoustic signals interpretable for the AI model, they must be pre-processed and converted into digital signals. Furthermore, the models need to receive data extracted from these signals as training features, so it is essential to use a feature extraction method.

One of the methods for extracting acoustic features is applying the MFCC algorithm, which involves a short-term analysis of the acoustic signals [43]. First, the acoustic signals are transformed using the Fourier transform. Next, triangular windows are applied to map the spectrum powers onto the Mel scale. After this, the logarithm of the powers is calculated and transformed using the discrete cosine transform, and the Mel frequencies are also computed. Finally, the feature vectors are obtained as the amplitudes of the spectrums [44].

The first step in extracting MFCC involves segmenting the acoustic signals into overlapping frames, as illustrated in Figure 3. Here, N denotes the length of each frame, and each frame overlaps with its neighboring frames by a length of N-M. By applying a window function, such as the Hamm or Hann window (as shown in Equations (1) and (2)), to the segmented and overlapped frames, the discontinuities in the resulting signal frames are minimized. Consequently, discrete windows are produced in the time domain, which can subsequently be transformed into the frequency domain using the Fourier transform, resulting in $S(k)$, which is the result of the frame with N samples [45].

$$(\text{Hamming Window})\ w(n) = 0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right),\ 0 \le n \le N-1 \tag{1}$$

$$(\text{Hanning Window})\ w(n) = 0.5 - 0.5cos\left(\frac{2\pi n}{N-1}\right),\ 0 \le n \le N-1 \tag{2}$$

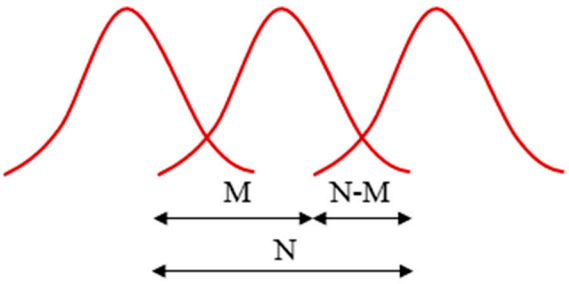


Figure 3. Schematic of windows and the segmenting of frames.

Mel Filter Banks convert the resulting spectra to the Mel scale. In fact, the linear scale frequencies (in Hz) will be mapped in the Mel scale by the Mel Bank which includes the triangular-shaped Mel scale filters. Equation (6) is the formula which gives Mel frequency (Mel(f)) when 'f' is the linear frequency (in Hz). Also, S_{mel} illustrated in Equation (7) is the result of multiplying the filter banks matrix (W) by the power spectrum (P). The log of the Mel scale spectrum is transformed by the discrete cosine transform, resulting in MFCCs without complex numbers. The discrete cosine transforms also make it easier to train the AI models by creating uncorrelated features. These final procedures have been depicted in Equations (8) and (9). If g = 0, 1, 2, . . ., J-1 (J is the number of MFCC), the C_g will be the g^{th} vector of MFCC and $S_{Mel}(m)$ will be the m^{th} Mel filter if the number of Mel filters in the filter bank is N_f. Usually, many MFCC are not needed; the first few coefficients are sufficient to represent the input signal [45].

$$S(k) = \sum_{n=0}^{N-1} s(n)e^{-j2\pi nk/N},\ 0 \le n \le N-1 \tag{3}$$

$$N_2 = \frac{N}{2} \tag{4}$$

$$P = |S(1 : N_2)|^2 \tag{5}$$

$$Mel(f) = 2595 \times log_{10}\left(1 + \frac{f}{700}\right) \tag{6}$$

$$S_{mel} = W \times P \tag{7}$$

$$C_g = dct(log_{10}(S_{Mel})) \tag{8}$$

$$C_g = \sqrt{\frac{2}{N_f}} \sum_{m=0}^{N_f - 1} log_{10}(S_{Mel}(m)) cos\left[\frac{g\pi}{N_f}(m + 0.5)\right] \tag{9}$$

In the current investigation, the Librosa package [46] was used as the feature extraction tool to process the input acoustic signal and extract the MFCC. Although this package can process the input signal and extract MFCC by default, a closer examination reveals that there are several adjustable processing parameters. These include the number of MFCC (n_mfcc), the sample rate (sr), the number of samples between successive frames (hop_length), and the length of the FFT window (n_fft). Therefore, we have decided to optimize these parameters to ensure the highest classification scores. Overall, the values of current, Ton, rotational speed, and surface roughness were labeled from zero to thirteen and the aim was to detect and predict the value corresponding to them separately based on the given acoustic signal to the multi-label ANN model.

3. Results and Discussion

Given the numerous hyperparameters in an ANN model, there are many possible structures for training. Finding the optimal combination of these parameters often requires extensive trial-and-error. Methods such as Keras Tuner [47] can help us address this challenge. The tuner can identify the best combination of model hyperparameters based on the ranges outlined in Table 3. Figure 4 illustrates the overall trend of feature extraction and the ANN structure.

Table 3. The optimum hyper-parameters for the model structure.

Hyper-Parameters	Range	Optimum Hyper-Parameters
Number of Layers	1 to 4	4
Number of Neurons	4 to 64	64-48-44-24
Learning Rate	0.001 to 0.1	0.01

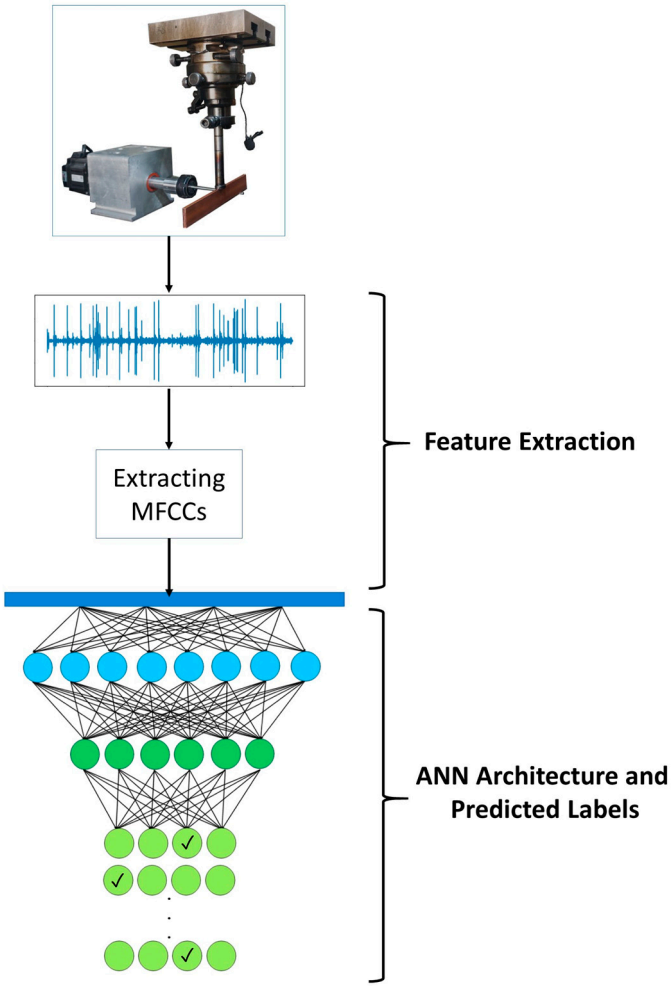


Figure 4. Schematic of feature extraction and the ANN architecture in which each hidden layer has a different color and each tick mark is the representative of the selected label.

As discussed, achieving optimal classification performance requires determining the best hyperparameters. To find the most effective preprocessing hyperparameters, we altered them systematically, and, ultimately, the highest predictive scores indicated the ideal hyperparameters. The sampling rate is one such hyperparameter that defines the frequency of extracting discrete values from a continuous signal. In the first step, the best model with the optimal sampling rate was selected and used to identify the best model with the most effective n_fft in the subsequent step. The researchers used a special value of preprocessing factors. For example, a 22,100 Hz or 20,000 Hz sample rate were applied in audio classifying [48] while another classifier used 22,050 Hz as its sample rate alongside the number of FFT of 2048 and a 512 hop length [49,50]. Also, the researchers usually use a number of MFCC between 10 and 20 [51]; however, 40 MFCC have also been used [52]. The value of the preprocessing factors highly influences the extracted information and, as such, the final trained model's performance. So, in order to find probable better preprocessing values producing a more accurate trained classifier, we examined other preprocessing values which were not very far from the values studied by the previous researchers but were not exactly the same. Our process allowed us to systematically optimize hyperparameters, including sr (among 4500, 9000, and 18,000 Hz), n_fft (among 256, 512, and 1024), hop_length (among 128, 256, and 512), and n_mfcc (among 20, 40, and 80) in four distinct steps. To evaluate the performance of each model, three significant criteria were considered: precision, recall, and F1 score. Precision is calculated by dividing the number of test samples that have been correctly predicted as positive by the

total number of test samples predicted as positive (both true positives and false positives). Recall, on the other hand, measures the fraction of test samples that have been correctly predicted as positive out of the total number of actual positive samples. Both precision and recall values indicate the improved performance of our AI model when they have higher values; however, the highest possible value for these scores is 1. Additionally, the F1 score is defined as the weighted harmonic mean of precision and recall, providing an overall depiction of the model's performance [53]. As mentioned earlier, all the data (the features extracted from the acoustic signals) is divided into training, validation, and test sets. The model learns to recognize the true labels when it encounters any of the training data. Meanwhile, it evaluates its performance by testing its predictions on the validation data. If the model does not perform well on the validation data, it will adjust its predictions to improve. Typically, the loss should decrease while the accuracy should increase during the training process. Although achieving high accuracy on both the training and validation datasets suggests that the model has been well trained, our experience shows that even if both the training and validation accuracies reach about their maximum value of 1 (or 100%), as illustrated in Figures 5 and 6, this does not guarantee high performance in predicting the test data, which has not been included in the model's training at all.

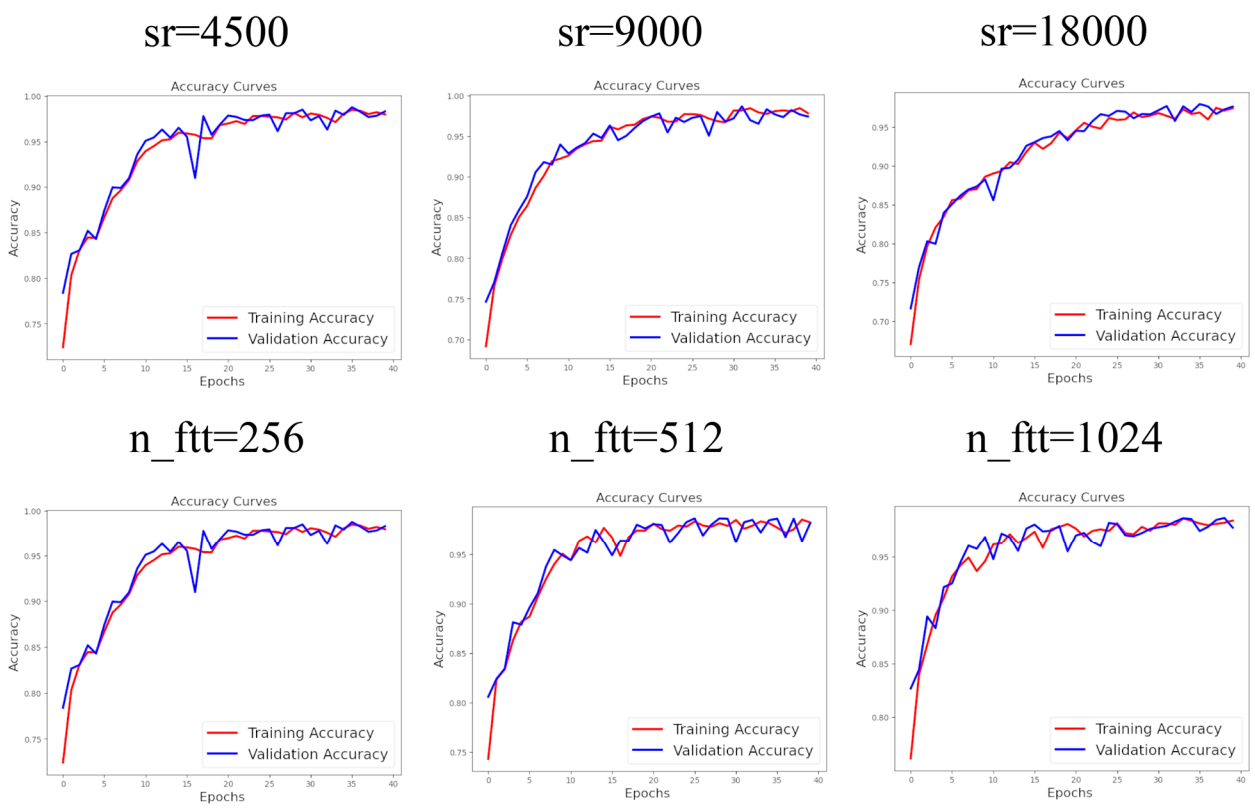


Figure 5. Accuracy diagrams for different sample rates (sr) and numbers of fft (n_fft).

It is evident from Figures 5 and 6 that some of the ANN models do not achieve maximum accuracy on the training and validation data during training. However, while some models do attain an accuracy of about 1 (or 100%), this does not guarantee that these models can correctly predict all labels of new, unseen data. Therefore, despite comparing the training accuracy diagrams in Figures 5 and 6, we selected the best hyperparameters based on the models' performance on the test data. As discussed, the models were evaluated using a test dataset that included 658 files.

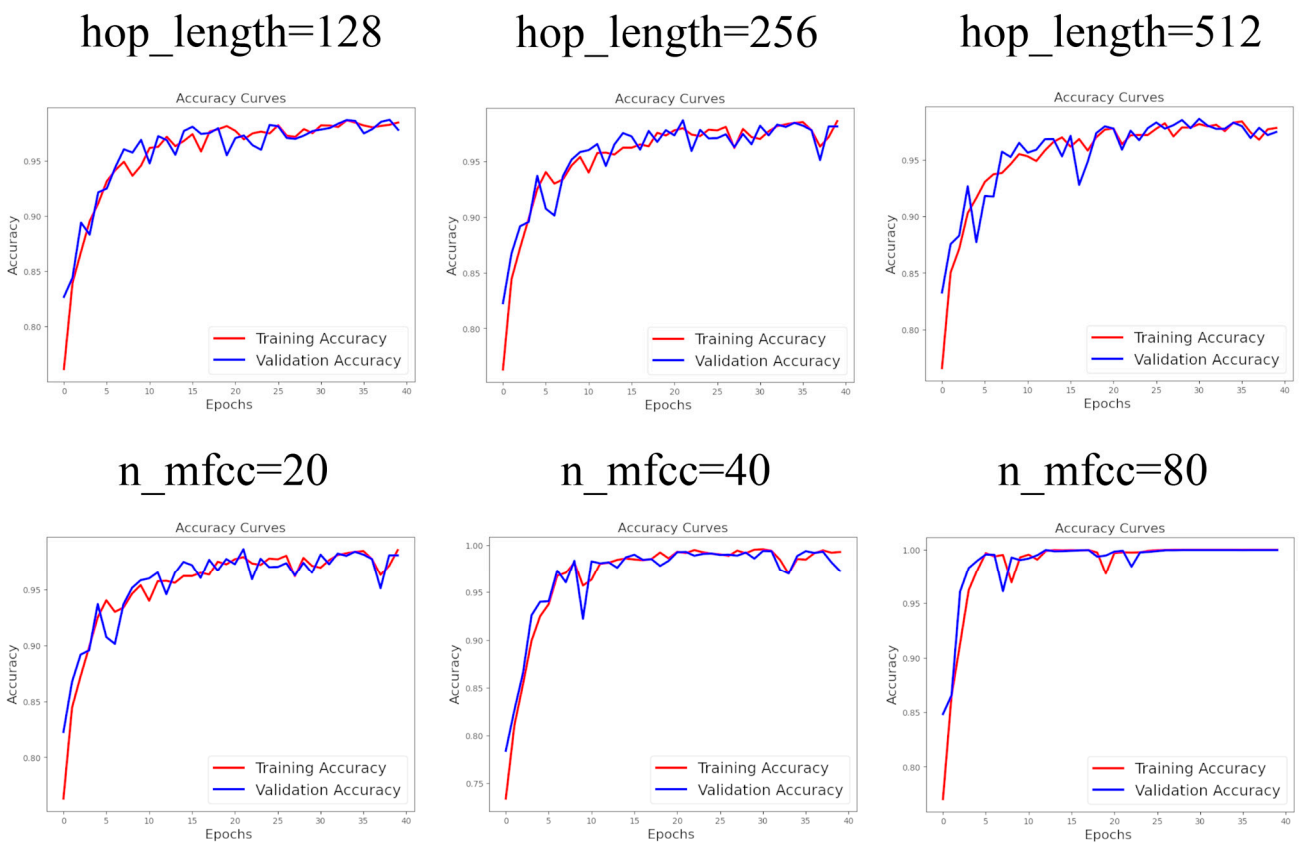


Figure 6. Accuracy diagrams for different numbers of MFCC (n_mfcc) and hop length.

In the first step, we modified the sample rate values during the preprocessing of the acoustic signals. After training three models with sample rates of 4500, 9000, and 18,000 Hz, respectively, we tested these models using the test data. The results are summarized in Table 4, which shows that the model with the lowest sample rate yielded the best results concerning the micro average, macro average, weighted average, and samples average. This model achieved 100%, 77%, 87% for the precision, recall, and F1 score, respectively. However, this model could not predict the labels 8, 10, and 11 well and some the prediction scores fell down to less than 10%. This shows that this model needs to be more investigated to increase its performance. The two other models trained with the data extracted with 9000 and 18,000 Hz showed a lower performance in comparison to the first model by 1 to 7% in the average scores. However, all the models exhibited poor scores for some labels (even below 10%), particularly in recall and F1 score, indicating the need for improvement through other hyperparameters. At first glance, it may seem that gathering more features from each signal by applying a higher sample rate would result in richer training data, leading to better training and prediction. However, the results showed the opposite trend. The reason for these results may be that the acoustic data are not very complex and training its behavior into the ANN models does not require highly extracted features. Overall, a sample rate of 4500 was chosen as the best sample rate for further testing.

Table 4. Classification report and the scores for 4500, 9000, and 18,000 Hz sample rates.

	Sample Rate: 4500 Hz			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.88	0.93	333
1	1.00	0.93	0.96	325
2	0.99	0.86	0.92	331
3	1.00	0.86	0.92	327
4	1.00	0.75	0.85	331
5	1.00	0.72	0.84	327
6	1.00	1.00	1.00	86
7	0.98	0.81	0.89	80
8	1.00	0.09	0.16	81
9	1.00	0.76	0.86	86
10	1.00	0.06	0.12	81
11	1.00	0.14	0.25	84
12	1.00	0.87	0.93	79
13	1.00	1.00	1.00	81
micro avg	1.00	0.77	0.87	2632
macro avg	1.00	0.69	0.76	2632
weighted avg	1.00	0.77	0.84	2632
samples avg	0.96	0.77	0.83	2632
	Sample Rate: 9000 Hz			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.92	0.96	333
1	0.99	0.96	0.98	325
2	1.00	0.83	0.91	331
3	1.00	0.91	0.95	327
4	1.00	0.74	0.85	331
5	1.00	0.52	0.68	327
6	1.00	1.00	1.00	86
7	0.99	0.86	0.92	80
8	1.00	0.01	0.02	81
9	1.00	0.90	0.94	86
10	1.00	0.05	0.09	81
11	0.00	0.00	0.00	84
12	1.00	0.75	0.86	79
13	1.00	0.99	0.99	81
micro avg	1.00	0.75	0.86	2632
macro avg	0.93	0.67	0.73	2632
weighted avg	0.97	0.75	0.82	2632
samples avg	0.97	0.75	0.82	2632
	Sample Rate: 18,000 Hz			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.98	0.99	333
1	1.00	0.90	0.95	325
2	1.00	0.77	0.87	331
3	0.99	0.83	0.91	327
4	1.00	0.68	0.81	331
5	0.99	0.58	0.73	327

Table 4. *Cont.*

	Sample Rate: 18,000 Hz			
	Precision	**Recall**	**F1-Score**	**Support**
6	1.00	1.00	1.00	86
7	1.00	0.80	0.89	80
8	1.00	0.43	0.60	81
9	1.00	0.67	0.81	86
10	1.00	0.01	0.02	81
11	1.00	0.05	0.09	84
12	0.99	0.90	0.94	79
13	0.96	1.00	0.98	81
micro avg	0.99	0.75	0.85	2632
macro avg	0.99	0.69	0.76	2632
weighted avg	0.99	0.75	0.82	2632
samples avg	0.98	0.75	0.81	2632

In the next step, the length of the Fast Fourier Transform window (n_fft) was considered with values of 256, 512, and 1024, and the prediction scores were compared to identify the best model based on these scores. Among the three tested models, the one with n_fft equal to 1024 achieved the highest scores across micro average, macro average, weighted average, and sample average on the test data, as shown in Table 5. Notably, this model maintained an average score of over 80% in all except the macro average of the recall which dropped to 78%. The two other models which had been trained with 256 and 512 n_fft resulted in poor results by up to 9% lower average scores in comparison to the model which had been trained with 1024 n_fft. All the three models had poor results in predicting the labels 8, 10, and 11. The model which used the 256 n_fft predicted the label 8 by 9% and 16% for recall and F1 score. This model also predicted the label 10 by 6% and 12% for recall and F1 score and label 11 by 14 and 25% for recall and F1 score. Moreover, the model which used the 512 n_fft predicted the label 8, 10, and 11 by 4% to 20% for recall and 7% to 34% for F1 score. However, the model which used the 1027 n_fft predicted the label 8, 10, and 11 by 10% to 64% for recall and 18% to 78% for F1 score.

The signal segments are transformed from time domain to frequency domain by the Fast Fourier Transform and the wider segment has more frequency information available. These results indicate that when capturing a segment of the signal to extract its features, increasing the length of the capturing window can lead to the extraction of more useful information from the signal for training the ANN model. This suggests that capturing more details with wider windows enhances feature extraction, providing better training data for each acoustic signal. Overall, although the model with n_fft equal to 1024 resulted in high prediction scores, there are still recall and F1 scores (for label 10) falling below 20%, which is unacceptable. Therefore, we aimed to improve these scores and investigate the effect of hop length on them in the next stage.

Changing the hop length from 128 to 512 did not significantly affect the average scores, as observed in the results presented in Table 6. However, increasing the hop length from 128 to 256 led to an improvement of 1% in some of the average recall and F1 scores. As shown in Table 6, after changing the hop length from 128 to 256, average precision, recall, and F1 scores changed from 100%, 83%, and 91% to 100%, 84%, and 91% for the micro average scores, respectively, and such a similar increase was observed in the other average scores. The low prediction scores can be observed again in the prediction of some labels like label 10 which has been predicted by the first model by 100%, 10%, and 18% for the precision, recall and, F1, respectively. This label has been predicted by the second model by 100%, 11%, and 20% for the precision, recall, and F1, respectively. The third model also had poor

results on the label 10 by 100%, 6%, and 12% for the precision, recall, and F1, respectively. These results showed that we have not arrived at an overall acceptable performance yet and the investigation should be continued to improve the prediction results.

Table 5. Classification report and the scores for 256, 512, and 1024 n_fft.

	n_fft: 256			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.88	0.93	333
1	1.00	0.93	0.96	325
2	0.99	0.86	0.92	331
3	1.00	0.86	0.92	327
4	1.00	0.75	0.85	331
5	1.00	0.72	0.84	327
6	1.00	1.00	1.00	86
7	0.98	0.81	0.89	80
8	1.00	0.09	0.16	81
9	1.00	0.76	0.86	86
10	1.00	0.06	0.12	81
11	1.00	0.14	0.25	84
12	1.00	0.87	0.93	79
13	1.00	1.00	1.00	81
micro avg	1.00	0.77	0.87	2632
macro avg	1.00	0.69	0.76	2632
weighted avg	1.00	0.77	0.84	2632
samples avg	0.96	0.77	0.83	2632
	n_fft: 512			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.87	0.93	333
1	1.00	0.84	0.91	325
2	1.00	0.76	0.87	331
3	1.00	0.85	0.92	327
4	1.00	0.80	0.89	331
5	1.00	0.72	0.83	327
6	1.00	1.00	1.00	86
7	1.00	0.79	0.88	80
8	1.00	0.14	0.24	81
9	0.99	0.92	0.95	86
10	1.00	0.04	0.07	81
11	1.00	0.20	0.34	84
12	1.00	0.92	0.96	79
13	1.00	0.99	0.99	81
micro avg	1.00	0.76	0.86	2632
macro avg	1.00	0.70	0.77	2632
weighted avg	1.00	0.76	0.84	2632
samples avg	0.92	0.76	0.81	2632
	n_fft: 1024			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.91	0.95	333
1	1.00	0.94	0.97	325
2	1.00	0.95	0.97	331
3	1.00	0.89	0.94	327
4	1.00	0.84	0.91	331

Table 5. *Cont.*

	n_fft: 1024			
	Precision	**Recall**	**F1-Score**	**Support**
5	1.00	0.66	0.80	327
6	1.00	1.00	1.00	86
7	1.00	0.94	0.97	80
8	0.98	0.52	0.68	81
9	1.00	0.80	0.89	86
10	1.00	0.10	0.18	81
11	1.00	0.64	0.78	84
12	1.00	0.77	0.87	79
13	1.00	0.98	0.99	81
micro avg	1.00	0.83	0.91	2632
macro avg	1.00	0.78	0.85	2632
weighted avg	1.00	0.83	0.89	2632
samples avg	0.96	0.83	0.87	2632

Table 6. Classification report and the scores for 128, 256, and 512 hop length.

	Hop Length: 128			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.91	0.95	333
1	1.00	0.94	0.97	325
2	1.00	0.95	0.97	331
3	1.00	0.89	0.94	327
4	1.00	0.84	0.91	331
5	1.00	0.66	0.80	327
6	1.00	1.00	1.00	86
7	1.00	0.94	0.97	80
8	0.98	0.52	0.68	81
9	1.00	0.80	0.89	86
10	1.00	0.10	0.18	81
11	1.00	0.64	0.78	84
12	1.00	0.77	0.87	79
13	1.00	0.98	0.99	81
micro avg	1.00	0.83	0.91	2632
macro avg	1.00	0.78	0.85	2632
weighted avg	1.00	0.83	0.89	2632
samples avg	0.96	0.83	0.87	2632
	Hop Length: 256			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.92	0.96	333
1	1.00	0.92	0.96	325
2	1.00	0.91	0.95	331
3	1.00	0.94	0.97	327
4	1.00	0.82	0.90	331
5	1.00	0.76	0.86	327
6	1.00	1.00	1.00	86
7	1.00	0.84	0.91	80

Table 6. *Cont.*

	Hop Length: 256			
	Precision	**Recall**	**F1-Score**	**Support**
8	1.00	0.57	0.72	81
9	0.99	0.87	0.93	86
10	1.00	0.11	0.20	81
11	1.00	0.46	0.63	84
12	1.00	0.89	0.94	79
13	1.00	0.99	0.99	81
micro avg	1.00	0.84	0.91	2632
macro avg	1.00	0.79	0.85	2632
weighted avg	1.00	0.84	0.90	2632
samples avg	0.97	0.84	0.88	2632
	Hop Length: 512			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.85	0.92	333
1	1.00	0.94	0.97	325
2	1.00	0.85	0.92	331
3	1.00	0.81	0.90	327
4	1.00	0.80	0.89	331
5	1.00	0.61	0.75	327
6	1.00	1.00	1.00	86
7	1.00	0.85	0.92	80
8	1.00	0.16	0.28	81
9	1.00	0.79	0.88	86
10	1.00	0.06	0.12	81
11	1.00	0.56	0.72	84
12	1.00	0.72	0.84	79
13	0.98	1.00	0.99	81
micro avg	1.00	0.77	0.87	2632
macro avg	1.00	0.71	.0.79	2632
weighted avg	1.00	0.77	0.85	2632
samples avg	0.94	0.77	0.82	2632

The limited impact of different hop length values may be attributed to the fact that the hop length defines the number of samples between successive frames. Changing the hop length only changes the relative positions between the consecutive frames and the frames might become closer to or further from each other. When they are closer to each other, the common part between them is further apart. This means that when using smaller hop lengths, there is more information that is extracted by both consecutive frames. Because the AI model does not need to consider this information twice, it might seem that using higher hop lengths (resulting in less repetitive extracted data) could be preferable, but as our results show, setting the hop length in a very high value (512) has decreased the recall and F1 score from 84% and 91% to 77% and 87%, respectively. This negative effect is in relation with the windowing function effect. If the windowing function has zero or near zero values at its start and end part, when applying it to the segment, it will lead to a loss of some of the signal information. To address this challenge, using an overlap between each two consecutive frames can decrease this effect. So, when very high values of hop length are applied between two consecutive frames, the common part between them is decreased and the negative effect of the windowing function may happen, resulting in the loss of some important information, which may affect the trained model's performance. Considering both the positive and negative aspects of applying the high and low values of hop length, selecting an optimum value is suggested, which, in our case, was 256.

Increasing the number of MFCC from 20 to 40 resulted in a significant improvement in recall and F1 score, particularly for label 10, which had been problematic for all previous models. The recall and F1 score for label 10 increased from 11% and 20% to 41% and 58%, respectively. Additionally, increasing the number of MFCC from 20 to 40 led to approximately a 2% improvement in average recall and F1 scores; however, the average precision declined by 1% to 2% across micro, macro, and weighted averages. As depicted in Table 7, the second model (with 40 MFCC) obtained 99%, 86%, and 92% for the precision, recall, and F1, respectively, considering the micro average scores. The rest of the model's scores were approximately close to the micro average scores with a 5% tolerance.

Remarkably, increasing the number of MFCC from 40 to 80 exceeded expectations, boosting the recall and F1 score for the label 10 from 41% and 58% to an impressive 100% for both metrics. This model, with 80 MFCC, also achieved 100% in all average scores, which is quite remarkable. When we look at the prediction scores for all the labels, we can see that the model has predicted them by obtaining the highest maximum precision, recall, and F1 score, which is 100%, and there are only some 99% scores for labels 3, 9, and 12. As described, the final result of obtaining MFCC from an acoustic signal is a set of coefficients which represent the behavior of the signal. Representing the acoustic signal behavior could be performed by an arbitrary number of coefficients based on the case study. In our case study, obtaining better results by training the model with larger values of MFCC in preprocessing procedure suggests that using MFCC with more coefficients provides a better representation of the acoustic signal and is more suitable for training the ANN model.

Table 7. Classification report and the scores for 20, 40, and 80 MFCC.

	mfcc: 20			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.92	0.96	333
1	1.00	0.92	0.96	325
2	1.00	0.91	0.95	331
3	1.00	0.94	0.97	327
4	1.00	0.82	0.90	331
5	1.00	0.76	0.86	327
6	1.00	1.00	1.00	86
7	1.00	0.84	0.91	80
8	1.00	0.57	0.72	81
9	0.99	0.87	0.93	86
10	1.00	0.11	0.20	81
11	1.00	0.46	0.63	84
12	1.00	0.89	0.94	79
13	1.00	0.99	0.99	81
micro avg	1.00	0.84	0.91	2632
macro avg	1.00	0.79	0.85	2632
weighted avg	1.00	0.84	0.90	2632
samples avg	0.97	0.84	0.88	2632
	mfcc: 40			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	1.00	1.00	333
1	1.00	0.98	0.99	325
2	1.00	0.90	0.95	331
3	0.99	0.99	0.99	327
4	1.00	0.60	0.75	331

Table 7. *Cont.*

	mfcc: 40			
	Precision	**Recall**	**F1-Score**	**Support**
5	0.95	0.88	0.91	327
6	1.00	1.00	1.00	86
7	1.00	0.81	0.90	80
8	0.92	0.98	0.95	81
9	0.97	0.45	0.62	86
10	1.00	0.41	0.58	81
11	1.00	0.52	0.69	84
12	0.93	1.00	0.96	79
13	1.00	0.81	0.90	81
micro avg	0.99	0.86	0.92	2632
macro avg	0.98	0.81	0.87	2632
weighted avg	0.99	0.86	0.90	2632
samples avg	0.98	0.86	0.89	2632
	mfcc: 80			
	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	1.00	1.00	333
1	1.00	1.00	1.00	325
2	1.00	1.00	1.00	331
3	1.00	0.99	1.00	327
4	1.00	1.00	1.00	331
5	1.00	1.00	1.00	327
6	1.00	1.00	1.00	86
7	1.00	1.00	1.00	80
8	1.00	1.00	1.00	81
9	1.00	0.99	0.99	86
10	1.00	1.00	1.00	81
11	1.00	1.00	1.00	84
12	1.00	0.99	0.99	79
13	1.00	1.00	1.00	81
micro avg	1.00	1.00	1.00	2632
macro avg	1.00	1.00	1.00	2632
weighted avg	1.00	1.00	1.00	2632
samples avg	1.00	1.00	1.00	2632

4. Conclusions

AI models have found their place in various fields, including manufacturing. However, these models face certain challenges, such as the need for significant time and cost investments to prepare sufficient training data. To address this issue, we propose using acoustic-based data as training data. This type of data are not only easy to record with a simple microphone but also contains valuable information about process behavior. This way, we could solve the problem by gathering enough training data. This method of gathering data does not have high costs in comparison to other methods like the acoustic emission method since it needs only a microphone to receive the generated sounds.

In this study, the hybrid EDT process was considered as a case study. It is a complex process with multiple influencing parameters. However, there were concerns regarding the suitability of these data for training an accurate model. Therefore, we decided to evaluate this capability, especially since the models are expected to predict various labels, such as surface roughness, Tons, currents, and rotational speeds, separately using ANNs for multilabel classification, rather than using typical classification models which had been used in the previous studies. After recording the sounds generated during EDT

under different Tons, currents, and rotational speeds, the recorded sounds were used as training data for the ANN models. Since acoustic data needs to be preprocessed and its features extracted before being fed as input into the ANN models, we utilized the Librosa package for preprocessing. Due to the varying parameters involved in preprocessing—such as sample rate, FFT window length, hop length, and the number of MFCC—which can significantly impact the prediction results, we aimed to identify the optimal preprocessing parameters for the current dataset. While many researchers may prefer using the Python packages in default setting, our results proved that it is necessary to delve into preprocessing parameters to find their best values resulting in the highest possible performance for the trained model. The final trained AI model with the highest prediction performance can be used in surface roughness evaluation as an alternative method or it can be used to online monitoring of the EDT process by developing a mobile app which has a microphone to receive the generated acoustic signals. Furthermore, the authors believe that the obtained results can be expanded for other manufacturing processes applying some adjustments in the collected data and the programming codes. Our approach considered maximizing precision, recall, and F1 scores across four steps, and the results are presented as follows:

1. Among the three different sampling rates tested (4500 Hz, 9000 Hz, and 18,000 Hz), applying a sample rate of 4500 Hz resulted in better training outcomes, with all average prediction scores exceeding 69%. This performance surpassed that of the two other ANN models. These results indicate that increasing the rate of data collection from recorded sounds does not necessarily lead to more effective training data.
2. Changing the length of the FFT windows significantly affects the prediction scores. Among the three considered values of n_fft (256, 512, and 1024), the model using the highest n_fft yielded the best prediction scores. All average scores in precision, recall, and F1 score improved by more than 9% compared to the best average scores from the previous models. These results demonstrate that capturing more details with wider windows can enhance the feature extraction of the acoustic signal used for training data.
3. The value of hop length did not significantly affect the training and prediction results. However, increasing the hop length from 128 to 256 led to a 1% increase in some average recall and F1 scores. Therefore, the number of samples between successive frames, ranging from 128 to 512, did not have a remarkable effect on the prediction scores.
4. Increasing the number of MFCC in preprocessing significantly affected the prediction results. While increasing this parameter from 20 to 40 improved the average recall and F1 scores by only 2%, raising it from 40 to 80 led to a substantial enhancement in the prediction scores. The corresponding ANN model achieved maximum average scores of 100%. This model demonstrated exemplary performance, even when predicting labels such as label 10, for which the best recall and F1 scores in all previously trained models were only 41% and 58%, respectively.

Author Contributions: Conceptualization M.S. and M.H.; methodology: M.S. and M.H.; software: M.S.; visualization: M.S. and M.H.; validation: M.S.; formal analysis: M.S.; investigation: M.S.; resources: M.S.; data curation: M.S.; writing—original draft preparation: M.S.; writing—review and editing: M.S. and M.H.; supervision: M.H.; project administration: M.H.; funding acquisition: M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are unavailable due to privacy restrictions.

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

References

1. Pruncu, C.I.; Zbitou, J. *Advanced Manufacturing Methods: Smart Processes and Modeling for Optimization*; CRC Press: Boca Raton, FL, USA, 2022; ISBN 1000643581.
2. Liao, Y.; Ragai, I.; Huang, Z.; Kerner, S. Manufacturing Process Monitoring Using Time-Frequency Representation and Transfer Learning of Deep Neural Networks. *J. Manuf. Process.* **2021**, *68*, 231–248. [CrossRef]
3. Mekhalfa, F.; Nacereddine, N. Multiclass Classification of Weld Defects in Radiographic Images Based on Support Vector Machines. In Proceedings of the 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, Morocco, 23–27 November 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–6.
4. Sudhagar, S.; Sakthivel, M.; Ganeshkumar, P. Monitoring of Friction Stir Welding Based on Vision System Coupled with Machine Learning Algorithm. *Measurement* **2019**, *144*, 135–143. [CrossRef]
5. Satpathy, M.P.; Mishra, S.B.; Sahoo, S.K. Ultrasonic Spot Welding of Aluminum-Copper Dissimilar Metals: A Study on Joint Strength by Experimentation and Machine Learning Techniques. *J. Manuf. Process.* **2018**, *33*, 96–110. [CrossRef]
6. Stadter, C.; Schmoeller, M.; von Rhein, L.; Zaeh, M.F. Real-Time Prediction of Quality Characteristics in Laser Beam Welding Using Optical Coherence Tomography and Machine Learning. *J. Laser Appl.* **2020**, *32*, 22046. [CrossRef]
7. Li, Y.; Yu, B.; Wang, B.; Lee, T.H.; Banu, M. Online Quality Inspection of Ultrasonic Composite Welding by Combining Artificial Intelligence Technologies with Welding Process Signatures. *Mater. Des.* **2020**, *194*, 108912. [CrossRef]
8. Amiri, N.; Farrahi, G.H.; Kashyzadeh, K.R.; Chizari, M. Applications of Ultrasonic Testing and Machine Learning Methods to Predict the Static & Fatigue Behavior of Spot-Welded Joints. *J. Manuf. Process.* **2020**, *52*, 26–34.
9. Jenab, A.; Sarraf, I.S.; Green, D.E.; Rahmaan, T.; Worswick, M.J. The Use of Genetic Algorithm and Neural Network to Predict Rate-Dependent Tensile Flow Behaviour of AA5182-O Sheets. *Mater. Des.* **2016**, *94*, 262–273. [CrossRef]
10. Dobrzański, L.A.; Maniara, R.; Sokolowski, J.H.; Krupiński, M. Modeling of Mechanical Properties of Al-Si-Cu Cast Alloys Using the Neural Network. *J. Achiev. Mater. Manuf. Eng.* **2007**, *20*, 347–350.
11. Abd El-Rehim, A.F.; Habashy, D.M.; Zahran, H.Y.; Soliman, H.N. Mathematical Modelling of Vickers Hardness of Sn-9Zn-Cu Solder Alloys Using an Artificial Neural Network. *Met. Mater. Int.* **2021**, *27*, 4084–4096. [CrossRef]
12. El-Rehim, A.; Alaa, F.; Zahran, H.Y.; Habashy, D.M.; Al-Masoud, H.M. Simulation and Prediction of the Vickers Hardness of AZ91 Magnesium Alloy Using Artificial Neural Network Model. *Crystals* **2020**, *10*, 290. [CrossRef]
13. Magalhães, F.C.; Ventura, C.E.H.; Abrão, A.M.; Denkena, B.; Breidenstein, B.; Meyer, K. Prediction of Surface Residual Stress and Hardness Induced by Ball Burnishing through Neural Networks. *Int. J. Manuf. Res.* **2019**, *14*, 295–310. [CrossRef]
14. Maleki, E.; Unal, O. Shot Peening Process Effects on Metallurgical and Mechanical Properties of 316 L Steel via: Experimental and Neural Network Modeling. *Met. Mater. Int.* **2021**, *27*, 262–276. [CrossRef]
15. Sembiring, J.; Amanov, A.; Pyun, Y.S. Artificial Neural Network-Based Prediction Model of Residual Stress and Hardness of Nickel-Based Alloys for UNSM Parameters Optimization. *Mater. Today Commun.* **2020**, *25*, 101391. [CrossRef]
16. Tuncer, T.; Akbal, E.; Aydemir, E.; Belhaouari, S.B.; Dogan, S. A Novel Local Feature Generation Technique Based Sound Classification Method for Covid-19 Detection Using Lung Breathing Sound. *Eur. J. Tech.* **2021**, *11*, 165–174. [CrossRef]
17. Bahrami, S.; Doraisamy, S.; Azman, A.; Nasharuddin, N.A.; Yue, S. CNN Architectures for Road Surface Wetness Classification from Acoustic Signals. In Proceedings of the Proceedings of the 8th International Conference on Computational Science and Technology, Toronto, ON, Canada, 24–25 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 777–788.
18. Gabrielli, L.; Ambrosini, L.; Vesperini, F.; Bruschi, V.; Squartini, S.; Cattani, L. Processing Acoustic Data with Siamese Neural Networks for Enhanced Road Roughness Classification. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.
19. Ramos-Romero, C.; León-Ríos, P.; Al-Hadithi, B.M.; Sigcha, L.; De Arcas, G.; Asensio, C. Identification and Mapping of Asphalt Surface Deterioration by Tyre-Pavement Interaction Noise Measurement. *Measurement* **2019**, *146*, 718–727. [CrossRef]
20. Gagliardi, A.; Staderini, V.; Saponara, S. An Embedded System for Acoustic Data Processing and AI-Based Real-Time Classification for Road Surface Analysis. *IEEE Access* **2022**, *10*, 63073–63084. [CrossRef]
21. Zhang, Z.; Wen, G.; Chen, S. Audible Sound-Based Intelligent Evaluation for Aluminum Alloy in Robotic Pulsed GTAW: Mechanism, Feature Selection, and Defect Detection. *IEEE Trans. Ind. Inform.* **2017**, *14*, 2973–2983. [CrossRef]
22. Pernambuco, B.S.G.; Steffens, C.R.; Pereira, J.R.; Werhli, A.V.; Azzolin, R.Z.; da Silva Diaz Estrada, E. Online Sound Based Arc-Welding Defect Detection Using Artificial Neural Networks. In Proceedings of the 2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE), Rio Grande, Rio Grande do Sul, Brazil, 23–25 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 263–268.
23. Lv, N.; Xu, Y.; Li, S.; Yu, X.; Chen, S. Automated Control of Welding Penetration Based on Audio Sensing Technology. *J. Mater. Process. Technol.* **2017**, *250*, 81–98. [CrossRef]
24. Kothuru, A.; Nooka, S.P.; Liu, R. Application of Audible Sound Signals for Tool Wear Monitoring Using Machine Learning Techniques in End Milling. *Int. J. Adv. Manuf. Technol.* **2018**, *95*, 3797–3808. [CrossRef]

25. Liu, M.-K.; Tseng, Y.-H.; Tran, M.-Q. Tool Wear Monitoring and Prediction Based on Sound Signal. *Int. J. Adv. Manuf. Technol.* **2019**, *103*, 3361–3373. [CrossRef]
26. Zhou, Y.; Sun, B.; Sun, W.; Lei, Z. Tool Wear Condition Monitoring Based on a Two-Layer Angle Kernel Extreme Learning Machine Using Sound Sensor for Milling Process. *J. Intell. Manuf.* **2022**, *33*, 247–258. [CrossRef]
27. Ibarra-Zarate, D.; Alonso-Valerdi, L.M.; Chuya-Sumba, J.; Velarde-Valdez, S.; Siller, H.R. Prediction of Inconel 718 Roughness with Acoustic Emission Using Convolutional Neural Network Based Regression. *Int. J. Adv. Manuf. Technol.* **2019**, *105*, 1609–1621. [CrossRef]
28. Bhandari, B. Comparative Study of Popular Deep Learning Models for Machining Roughness Classification Using Sound and Force Signals. *Micromachines* **2021**, *12*, 1484. [CrossRef] [PubMed]
29. Lee, C.-H.; Jwo, J.-S.; Hsieh, H.-Y.; Lin, C.-S. An Intelligent System for Grinding Wheel Condition Monitoring Based on Machining Sound and Deep Learning. *IEEE Access* **2020**, *8*, 58279–58289. [CrossRef]
30. Samraj, A.; Ragupathi, P.; Nandhakumar, T. Quantification of Tool Wear Condition from the Sound Recorded During the Process Using Neural Network. In *ICT for Competitive Strategies*; CRC Press: Boca Raton, FL, USA, 2020; pp. 309–315, ISBN 1003052096.
31. Li, Z.; Liu, X.; Incecik, A.; Gupta, M.K.; Królczyk, G.M.; Gardoni, P. A Novel Ensemble Deep Learning Model for Cutting Tool Wear Monitoring Using Audio Sensors. *J. Manuf. Process.* **2022**, *79*, 233–249. [CrossRef]
32. Azhiri, R.B.; Bideskan, A.S.; Javidpour, F.; Tekiyeh, R.M. Study on Material Removal Rate, Surface Quality, and Residual Stress of AISI D2 Tool Steel in Electrical Discharge Machining in Presence of Ultrasonic Vibration Effect. *Int. J. Adv. Manuf. Technol.* **2019**, *101*, 2849–2860. [CrossRef]
33. Gohil, V.; Puri, Y.M. Statistical Analysis of Material Removal Rate and Surface Roughness in Electrical Discharge Turning of Titanium Alloy (Ti-6Al-4V). *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2018**, *232*, 1603–1614. [CrossRef]
34. Gohil, V.; Puri, Y.M. A Study on the Effect of Tool Electrode Thickness on MRR, and TWR in Electrical Discharge Turning Process. In *IOP Conference Series: Proceedings of the Materials Science and Engineering, Melbourne, Australia, 15–16 September 2018*; IOP Publishing: Bristol, UK, 2018; Volume 346, p. 12036.
35. Jadidi, A.; Azhiri, R.B.; Teimouri, R. Electrical Discharge Turning by Assistance of External Magnetic Field, Part I: Study of MRR and Dimensional Accuracy. *Int. J. Light. Mater. Manuf.* **2020**, *3*, 265–276. [CrossRef]
36. Phang, Y.M.; Asmelash, M.; Hamedon, Z.; Azhari, A. Investigation on Turning Operation Using Die Sinking EDM Process. *Mater. Today Proc.* **2021**, *46*, 1569–1573. [CrossRef]
37. Çakıroğlu, R.; Günay, M. Comprehensive Analysis of Material Removal Rate, Tool Wear and Surface Roughness in Electrical Discharge Turning of L2 Tool Steel. *J. Mater. Res. Technol.* **2020**, *9*, 7305–7317. [CrossRef]
38. Guu, Y.H.; Hocheng, H. Effects of Workpiece Rotation on Machinability during Electrical-Discharge Machining. *Mater. Manuf. Process.* **2001**, *16*, 91–101. [CrossRef]
39. Azhiri, R.B.; Jadidi, A.; Teimouri, R. Electrical Discharge Turning by Assistance of External Magnetic Field, Part II: Study of Surface Integrity. *Int. J. Light. Mater. Manuf.* **2020**, *3*, 305–315. [CrossRef]
40. Matoorian, P.; Sulaiman, S.; Ahmad, M. An Experimental Study for Optimization of Electrical Discharge Turning (EDT) Process. *J. Mater. Process. Technol.* **2008**, *204*, 350–356. [CrossRef]
41. Cheng, W.; Hüllermeier, E. Combining Instance-Based Learning and Logistic Regression for Multilabel Classification (Resubmission). In Proceedings of the LWA 2009-Workshop-Woche: Lernen-Wissen-Adaptivitat-Learning, Knowledge, and Adaptivity, Darmstadt, Germany, 21–23 September 2009; Citeseer: University Park, PA, USA, 2009; pp. 22–29.
42. Kim, J.; Lee, H.; Jeong, S.; Ahn, S.-H. Sound-Based Remote Real-Time Multi-Device Operational Monitoring System Using a Convolutional Neural Network (CNN). *J. Manuf. Syst.* **2021**, *58*, 431–441. [CrossRef]
43. Desai, N.; Dhameliya, K.; Desai, V. Feature Extraction and Classification Techniques for Speech Recognition: A Review. *Int. J. Emerg. Technol. Adv. Eng.* **2013**, *3*, 367–371.
44. Lokesh, S.; Devi, M.R. Speech Recognition System Using Enhanced Mel Frequency Cepstral Coefficient with Windowing and Framing Method. *Clust. Comput.* **2019**, *22*, 11669–11679. [CrossRef]
45. Siam, A.I.; Elazm, A.A.; El-Bahnasawy, N.A.; El Banby, G.M.; El-Samie, A.; Fathi, E. PPG-Based Human Identification Using Mel-Frequency Cepstral Coefficients and Neural Networks. *Multimed. Tools Appl.* **2021**, *80*, 26001–26019. [CrossRef]
46. Available online: https://librosa.org/doc/main/generated/librosa.feature.mfcc.html (accessed on 25 November 2024).
47. Keras Tuner. Available online: https://keras.io/keras_tuner/ (accessed on 25 November 2024).
48. Ramesh, B.D.C.; Vishnu, R.S. CNN and Sound Processing-Based Audio Classifier for Alarm Sound Detection. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 365–375.
49. Cheng, Y.-H.; Chang, P.-C.; Nguyen, D.-M.; Kuo, C.-N. Automatic Music Genre Classification Based on CRNN. *Eng. Lett.* **2020**, *29*, 36–48.
50. Chillara, S.; Kavitha, A.S.; Neginhal, S.A.; Haldia, S.; Vidyullatha, K.S. Music Genre Classification Using Machine Learning Algorithms: A Comparison. *Int. Res. J. Eng. Technol.* **2019**, *6*, 851–858.

51. Bhat, A.; Krishna, A.V.; Acharya, S. Analytical Comparison of Classification Models for Raga Identification in Carnatic Classical Instrumental Polyphonic Audio. *SN Comput. Sci.* **2020**, *1*, 339. [CrossRef]
52. Vuppalapati, J.S.; Kedaru, S.; Kedari, S.; Ilapakurti, A.; Vuppalapati, C. Artificial Intelligent (AI) Clinical Edge for Voice Disorder Detection. In Proceedings of the Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys), London, UK, 3–4 September 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 2, pp. 750–766.
53. Grigorev, A. *Mastering Java for Data Science*; Packt Publishing Ltd.: Birmingham, UK, 2017; ISBN 1785887394.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.