



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

Sound-based Remote Manufacturing Process  
Monitoring using Convolutional Neural  
Network (CNN)

합성곱 신경망을 이용한 음성기반  
원격 제조공정 모니터링 시스템

2021 년 2 월

서울대학교 대학원

기계항공공학부

김 지 수

# 합성곱 신경망을 이용한 음성기반 원격 제조공정 모니터링 시스템

Sound-based Remote  
Manufacturing Process Monitoring  
using Convolutional Neural Network (CNN)

지도교수 안 성 훈

이 논문을 공학박사 학위논문으로 제출함

2020 년 10 월


서울대학교 대학원


기계항공공학부


김 지 수


김지수 의 공학박사 학위논문을 인준함

2020 년 12 월

위원장 : 강 연준 (인) 

부위원장 : 안 성훈 (인) 

위원 : 차성원 (인) 

위원 : 양인순 (인) 

위원 : 홍영진 (인) 

# Abstract

## Sound-based Remote Manufacturing Process Monitoring using Convolutional Neural Network (CNN)

Jisoo Kim

Department of Mechanical Aerospace Engineering  
The Graduate School  
Seoul National University

Smart factory is the main keyword in the field of manufacturing processes about the fourth industrial revolution. To realize the smart factory, making all pieces of device into smart devices that are connected to the centralized system to enable a real-time exchange of information is essential. Sound can be efficient means to make devices as smart devices because sound can contain the status information of various devices simultaneously, and it can be recorded easily outside of a device using only a microphone. In this study, multi-device operation monitoring system by analyzing sound is developed. Mic arrays for acquiring the sound were installed at the

outside the devices and recorded the sounds from several devices simultaneously. By analyzing the recorded sound with log-mel spectrogram and Convolutional Neural Network (CNN), the system could detect the operational status of three devices with an accuracy of 71-92%. To improve the performance, virtual data set was created by composition of individual device operating sounds of different intensities. With this virtual data set, accuracy can be enhanced to 87% ~ 99% accuracy and, required sound data amount could be reduced. Developed system was applied successfully in monitoring experiments in two different environments: a workshop in which hand-operated device was used and a factory with a computer numerical control machine and verifying the performance.

**Keywords :** Multi-device Monitoring, Sound monitoring, Convolutional Neuron Network (CNN), Smart factory

**Student Number :** 2016-30178

# 목 차

Chap. 1 Introduction.....	1
1.1 Fourth industrial revolution .....	1
1.2 Smart factory and smart devices .....	4
1.3 Methods of device monitoring .....	7
1.4 Sound monitoring and Convolutional Neuron Network.....	12
Chap 2. System Modeling .....	17
2.1 Concept of Convolutional Neural Network (CNN) .....	17
2.2 Fourier transform.....	22
2.3 Log-mel spectrogram.....	26
2.4 Proposed architecture.....	29
2.5 Concept of monitoring system.....	32
2.6 Parallel and independent system.....	37
Chap 3. Development of the monitoring system.....	39
3.1 Hardware of monitoring system.....	39
3.2 Training with actual data .....	44
3.3 Performance evaluation of monitoring system.....	49
3.4 Enhancement of performance.....	58
3.5 Virtual data set .....	68
3.6 Measuring intensity of sound based on masking .....	75

Chap 4. Applying to real factory .....	82
4.1 Case – Workshop with hand–operated device .....	84
4.2 Case – Factory with a CNC machine .....	89
4.2 Case – Factory with aluminum casting process .....	97
Chap 5. Application .....	103
5.1 Sound–based manufacturing process monitoring .....	103
Chap 6. Conclusion.....	107
참고문헌 .....	108
초록.....	116

## 표 목차

Table 1 Specification of hardware/software.....	36
Table 2 Specification of mic array.....	40
Table 3 Monitoring target device .....	43
Table 4 Specification of dataset for data training .....	46
Table 5 Specification of standard data.....	51
Table 6 Recognition performance with different mic position .....	55
Table 7 Test result with various neuron network.....	65
Table 8 Specification of the virtual data set.....	74
Table 9 Detail information about monitoring .....	85
Table 10 Detail information about monitoring .....	90
Table 11 Detail information about monitoring .....	98



# 그림 목차

Figure 1 Concept of 1st, 2nd, 3rd and 4th industrial revolution .....	3
Figure 2 Concept of smart factory .....	4
Figure 3 Concept of Cyber Physical System (CPS) .....	5
Figure 4 Comparison of monitoring method.....	7
Figure 5 Monitoring using vision .....	8
Figure 6 Monitoring using thermo-sensor .....	9
Figure 7 Monitoring using power consumption.....	10
Figure 8 Monitoring using acoustic emission.....	11
Figure 9 Overlap and separation of waves .....	12
Figure 10 Concept of Convolutional Neuron Network .....	13
Figure 11 Prediction of the quality of AM products with operation acoustic emission during process.....	14
Figure 12 Classifying the human activities with sonar and CNN ..	15
Figure 13 Classifying the genre of music with CNN .....	15
Figure 14 Concept of dropout .....	19
Figure 15 Signal processing method for pattern classification .....	23
Figure 16 Concept of Fourier Transform.....	23
Figure 17 Relation between frequency and log-mel spectrum static .....	26

Figure 18 Filterbank for log–mel spectrogram.....	27
Figure 19 Proposed convolutional neural network architecture for operational monitoring. ....	31
Figure 20 Concept diagram of sound data conversion from 1D to 2D .....	34
Figure 21 Schematic of the monitoring system.....	35
Figure 22 Signals received and processed from the by mic array	41
Figure 23 Operating sounds from the target devices after processing by short–time Fourier transform and log–mel spectrogram.....	42
Figure 24 Experimental setup for acquiring sample data to train the monitoring system. ....	45
Figure 25 Results of operational monitoring.....	47
Figure 26 Experimental setup for acquiring the test data set for evaluating the monitoring system.....	50
Figure 27 Performance of the monitoring system based on real data .....	53
Figure 28 Performance test with different mic position.....	55
Figure 29 Bandsaw operation monitoring with different operation environment.....	57
Figure 30 Fourier transformed operating sound .....	59
Figure 31 Performance evaluation of monitoring system with	

different target frequency ranges.....	59
Figure 32 Monitoring accuracy with different frequency range ...	61
Figure 33 Estimated accuracy with various frequency range .....	62
Figure 34 Accuracy of bandsaw monitoring with various neuron network.....	65
Figure 35 STFT based data process and wavelet transform based data process .....	67
Figure 36 Recognition performance with wavelet transform .....	67
Figure 37 Schematic of the creation of a virtual data set for training .....	72
Figure 38 Performance evaluation of monitoring system based on virtual data set .....	74
Figure 39 Concept diagram of intensity measuring .....	74
Figure 40 Fabrication of mask using existed operation sound data .....	78
Figure 41 Applying mask with intensity value .....	78
Figure 42 Measure the intensity of sound with different distance .....	79
Figure 43 Measure the intensity of sound: step test .....	80
Figure 44 Measure the intensity of sound with different devices	81
Figure 45 Clearly predicted section / not clearly predicted section	

.....	83
Figure 46 Workshop with hand-operated device .....	85
Figure 47 Monitoring results in a workshop with hand-operated devices .....	87
Figure 48 Process monitoring result of workshop with hand-operated devices .....	88
Figure 49 Factory with a Computer Numerical Control (CNC) machine .....	90
Figure 50 Monitoring results at a factory with a computer numerical control (CNC) machine .....	92
Figure 51 Process monitoring result of a factory with a computer numerical control (CNC) machine .....	96
Figure 52 Factory with aluminum casting process.....	98
Figure 53 Monitoring results at a factory with aluminum casting process .....	100
Figure 54 Process monitoring result of a factory with a factory with aluminum casting process .....	102
Figure 55 Example of Sound-based manufacturing process monitoring system .....	104
Figure 56 Drone position detection by sound analyzing .....	106

# **Chap. 1 Introduction**

## **1.1 Fourth industrial revolution**

The most important keyword that has recently hit the world will be the Fourth Industrial Revolution. However, while the impact of these Fourth Industrial Revolution is felt by all, it is not really defensive to define the Fourth Industrial Revolution. In the era of people and animals as power sources, the first industrial revolution that moved machines to power sources, the first industrial revolution that allowed mechanized businesses to turn electricity into power sources, the second industrial revolution that included various internal combustion engines using oil, the use of various plastics, and the integration of various information and communication technologies with the development of computers and the onset of the Internet, and the revolution that took place, and the revolution that took place. It was also clear that the technology driving the change and its impact. However, the 4th Industrial Revolution is being carried out simultaneously in many areas, with the level of large technological advancements that led to the existing revolution. The improvement in the performance of artificial neural networks enabled by the development of computing power shows that machines can excel in

many areas of recognition and optimization, and the development of information communication technology has allowed all equipment to be attached to and controlled in real time, making it possible for factories, homes, and offices to move like an organism, beyond the time when people were exchanging information directly through a small number of communication devices. For space engineering, once considered only for technology flaunting, it embodies a wide range of functions, enabling it to provide location information quickly and accurately to all devices around the world through a more sophisticated and popular GPS system. Also, these changes are not being carried out independently, but are being combined to create greater synergy.

In the wake of the fourth industrial revolution, many countries, universities, institutions, and companies are trying to develop relevant technology. The most commonly recognized aspect of this industrial revolution is the advancement of Information and Communication Technology (ICT) [1, 2]. Improvement of 5G telecommunication, which enables the fast response and large bandwidth, and Internet of Things (IoT), which enables the installation of telecommunication functions in all devices, make the rapid and widespread gathering of information possible. Cloud computing and big data management technology make it possible to

store information. Significant improvements in computing power have led to Artificial Intelligence (AI) with high performance via efficient data analysis techniques, such as machine learning [3–8].

After all, the most important keyword of the Fourth Industrial Revolution can be seen as "Connected" [9].

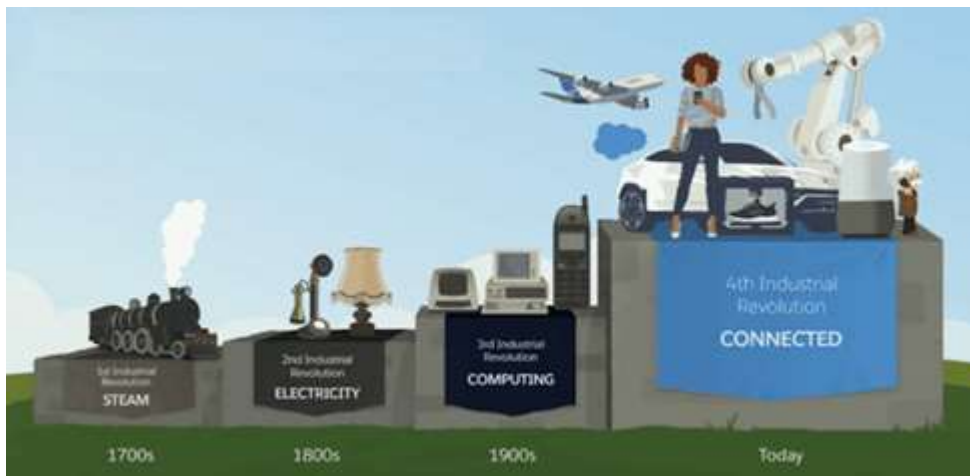


Figure 1 Concept of 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> industrial revolution [10]

## 1.2 Smart factory and smart devices

Smart factory is the main keyword in the field of manufacturing processes about the fourth industrial revolution [11]. Smart factory can be defined as manufacturing system with device connected to a cloud-based, centralized system and interactive information exchange functions through the internet and cloud. Through this centralized system, a higher level of monitoring, analysis, control, and design is possible, and thus smart factories are of interest as the future of manufacturing [12–15].

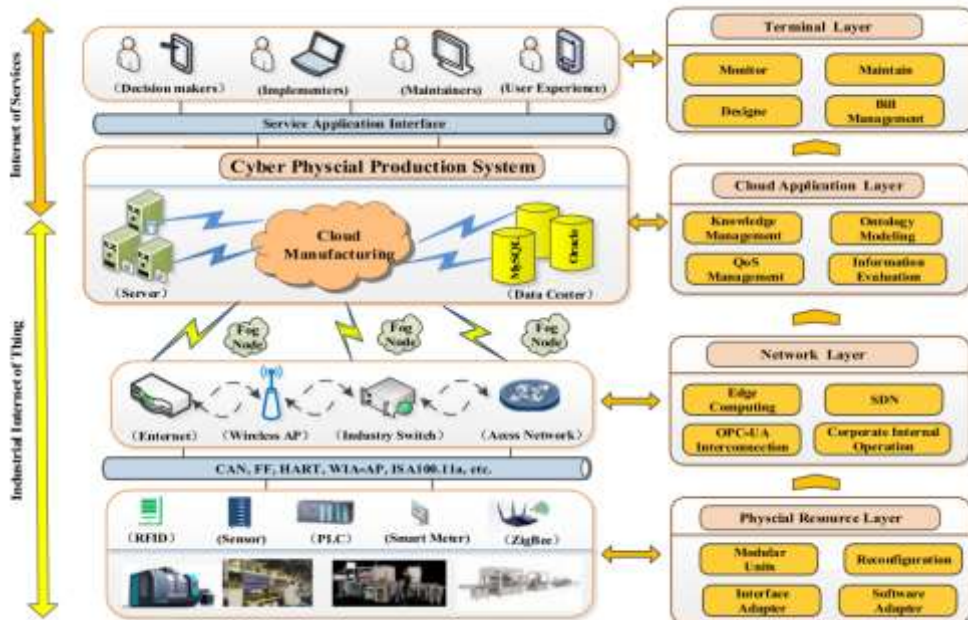


Figure 2 Concept of smart factory [16]



A Cyber-Physical System (CPS) is the virtual system that can simulate the manufacturing process based on the collected information. Currently, experimentation with device in a virtual space has gone beyond the level of individual pieces of device to test what happens in the factory in advance and to simulate the entire plant's performance [17]. Such CPSs have been applied to factory design and have begun to improve the performance of actual plants [18–29].

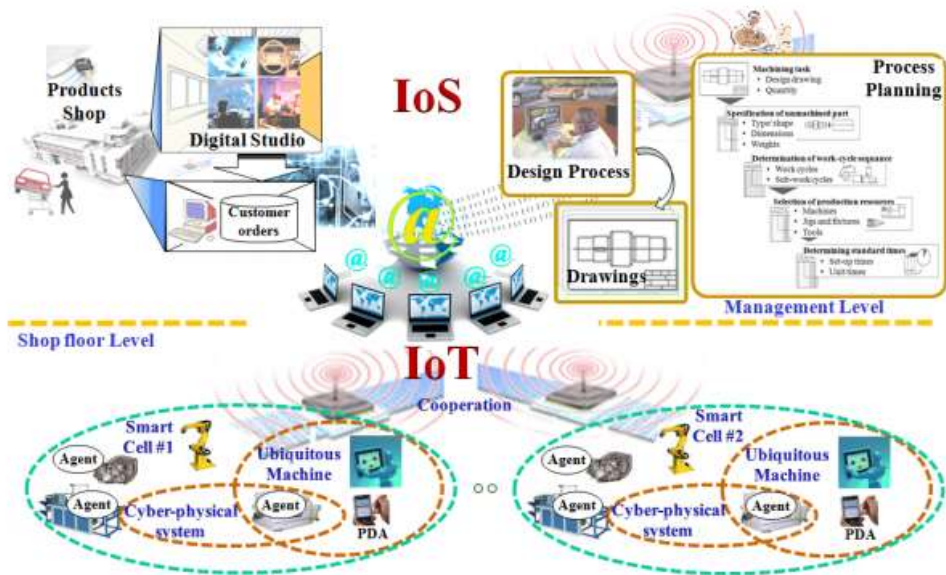



Figure 3 Concept of Cyber Physical System (CPS) [30]

The most important part of realizing a smart factory and a CPS is making all pieces of device into smart devices that connect to system based on the IoT. The status of each device can be shared in real time and gathered in the cloud to enable the better identification and control of device than was previously possible [23, 31–38]. Thus, it must be possible to collect detailed information about each device. Nowadays, newly developed device normally mount the IoT-based features to be smart device. However, device that was produced long ago typically lacks any means of connection to other device or systems. The technology to turn such device into smart devices easily and at a low cost is an urgent research task [39, 40]. This is more urgent for small companies which are likely to be more dependent on existing device than large enterprises that can design and build new plants and acquire new device [41]. Because small companies lack sufficient capital and are unable to renovate entire factories, it is difficult for them to introduce sweeping changes, which forces them to face the reality that they cannot match the pace of change elsewhere [42–44].

### 1.3 Methods of device monitoring

For these reasons, technologies are being developed to remotely monitor the status of a device using various methods, such as visual / sound / heat / power consumption [45].



Target Information	Vision / Visible Light	Sound / Acoustic Emission	Heat: Amount	Heat: Distribution	Power Consumption
Installation of Device	+	+	+	++	+++
Data Separation	+	++	+++	++	+
Data Overlap	+++	+	++	++	++
Block by Obstacles	+++	++	+++	+++	+
Analysis Device	+	+	+	+++	++
Digitize	+++	++	+	++	+
Size of Data	+++	+	+	+++	+

(+++ : Difficult / ++ : Normal / + : Easy)

Target Information	Vision / Visible Light	Sound / Acoustic Emission	Heat: Amount	Heat: Distribution	Power Consumption
Installation of Device	X	X	X	X	O
Data Separation	Easy	Difficult	Difficult	Easy	Easy
Data Overlap	Fully Covered	Fully Overlapped	Mixed	Mixed	Mixed
Block by Obstacles	Fully	Partially	Strongly	Strongly	N/A
Analysis Device	Eye / Camera	Microphone	Thermometer	Thermal Image Camera	Watt-Hour Meter
Digitize	Difficult	Difficult	Easy	Easy	Easy
Size of Data	3.5Mbps (MP4, 720p, 60hz)	32kbps (16kHz, wav)	20bps (10hz, raw data)	2.5Mbps (MP4, 720p, 30hz)	20bps (10hz, raw data)

Figure 4 Comparison of monitoring method

Visual methods are commonly used to observe the operation of device and to identify abnormal conditions [46–48]. Analyses of these methods have recently been enhanced by the use of Artificial Intelligence (AI) such as Convolutional Neuron Network (CNN) [49, 50]. Attempts have been made to read the information from the display panel installed at the device [42] or to visually recognize and analyze information maps of entire manufacturing process systems [51].

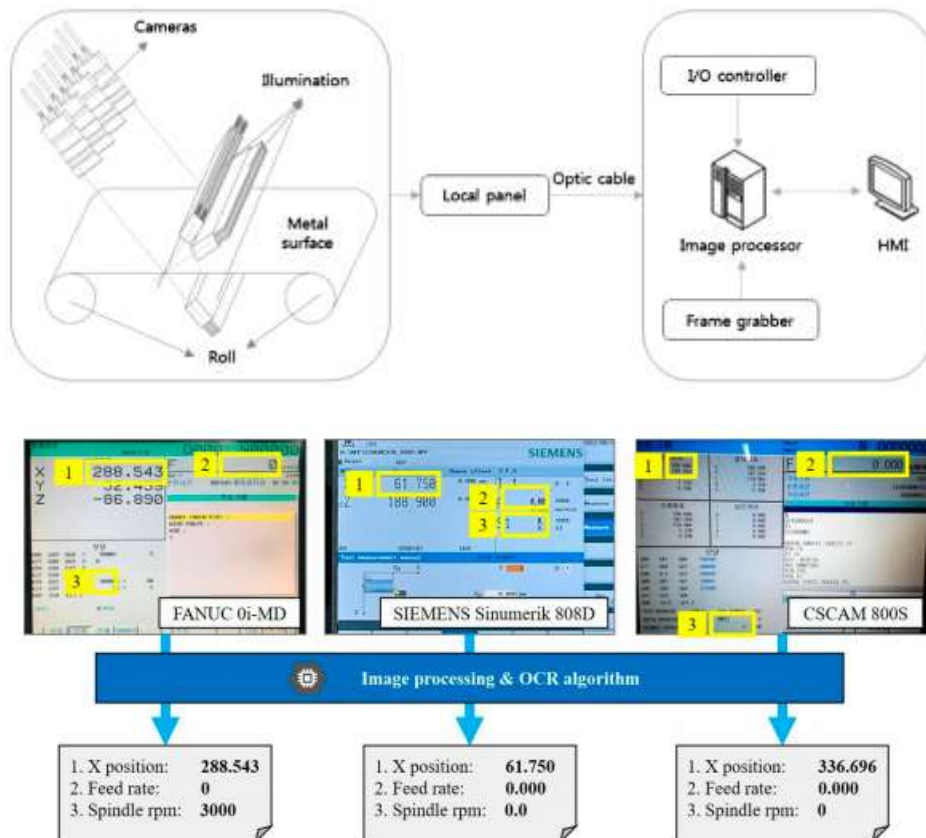


Figure 5 Monitoring using vision [42, 49]

Heat is the most basic piece of information used to understand the status of numerous pieces of device and plants, and controlling heating and cooling is a basic process in managing device [52]. Nowadays, various studies are being conducted using technology of measuring the distribution of heat, such as thermal imaging cameras which has improved significantly enough to make it possible to obtain information in real time [53, 54].

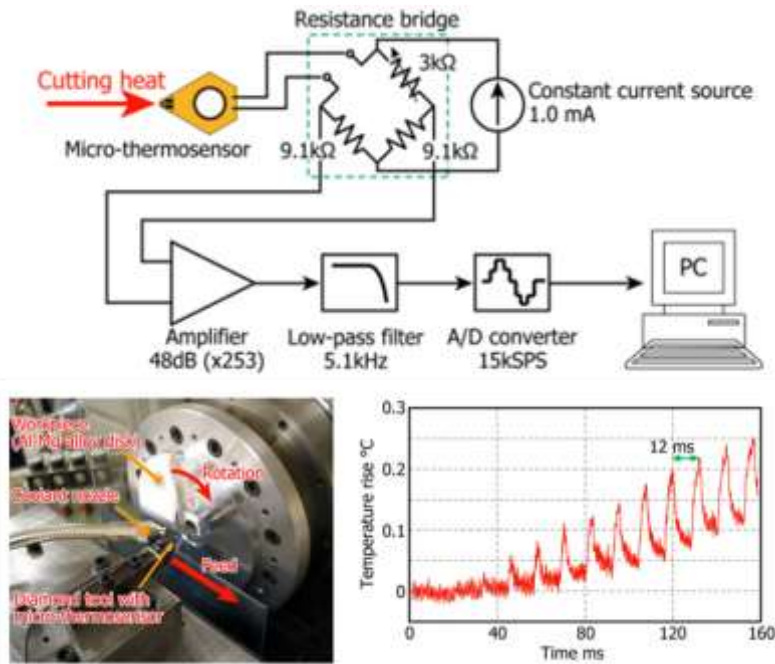


Figure 6 Monitoring using thermo-sensor [52]

Measuring power consumption requires the installation of an additional device, but studies are being conducted on this method because information about the target device can be extracted without requiring supplementary information and the data are highly reliable [55–57]

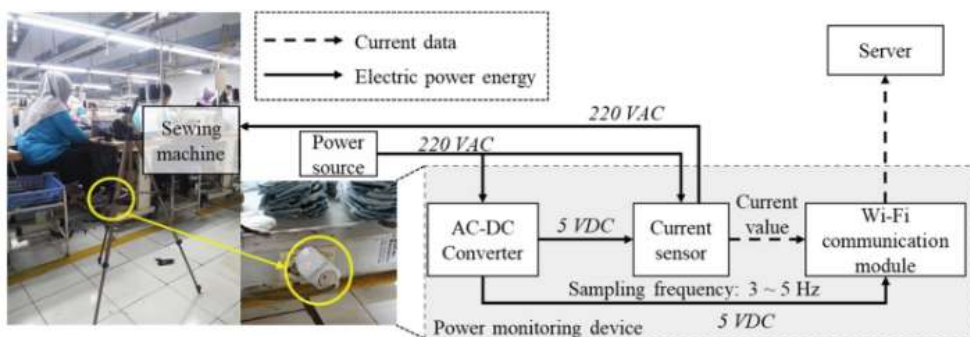
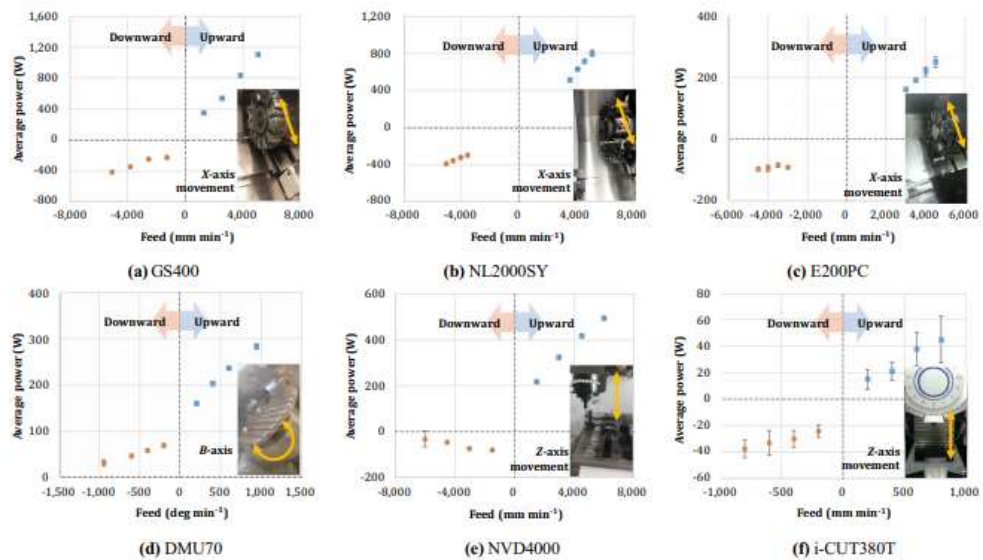


Figure 7 Monitoring using power consumption [56, 57]

Finally, sound is commonly used to identify the status of device. Because it shows good performance for diagnosing problems with mechanical parts, such as tool wear and vibration, many studies have been conducted [58, 59]. In addition, analyses of vibrations have been gaining attention recently, in which vibrations have been converted into two-dimensional (2D) data and used to classify the condition of machines with an AI tool used in image processing. The following section gives more details

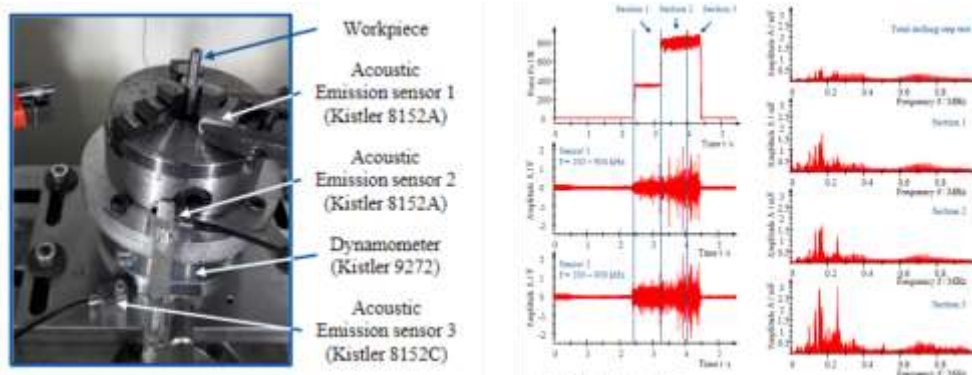


Figure 8 Manufacturing monitoring using acoustic emission [33]

## 1.4 Sound monitoring and Convolutional Neuron Network (CNN)

Sound is a vibration transmitted through a medium such as a gas, liquid, or solid. One of the notable characteristics of sound is that if various signals coincide, they overlap without affecting one another. Various pieces of information can be contained in sound, and this information is maintained even if there is interference from other factors. Thus, information can be obtained even when the surrounding environment is not controlled, and information about multiple sources can be acquired simultaneously.

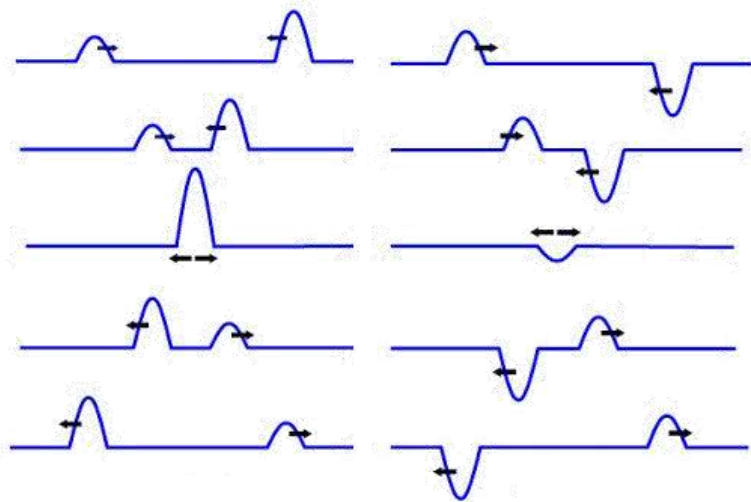


Figure 9 Overlap and separation of waves



Because processing sound produces a lot of data, recent analyses have used Artificial Neural Networks (ANN) [60, 61] or Support Vector Machines (SVM) [62], or Random Forest (RF) [63] rather than traditional methods of analysis. Analyzing sound via its one-dimensional raw signal is difficult, but it can be accomplished more easily if the sound is converted into 2D data via a Fourier transform and sorted by frequency. In the case of 2D data, studies are being conducted to classify signals in various ways. For instance, it is straightforward to apply Convolutional Neuron Network (CNN), which is frequently used in image processing [64, 65].

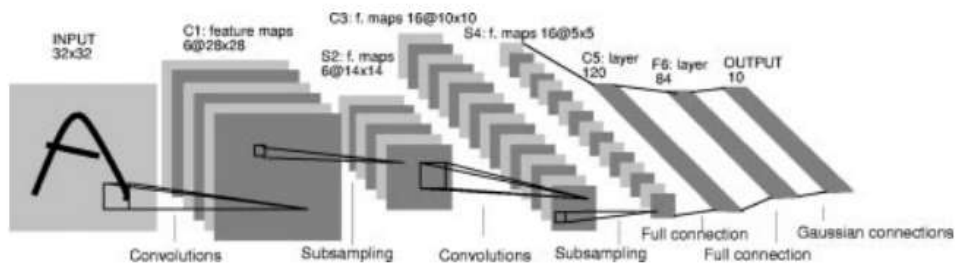
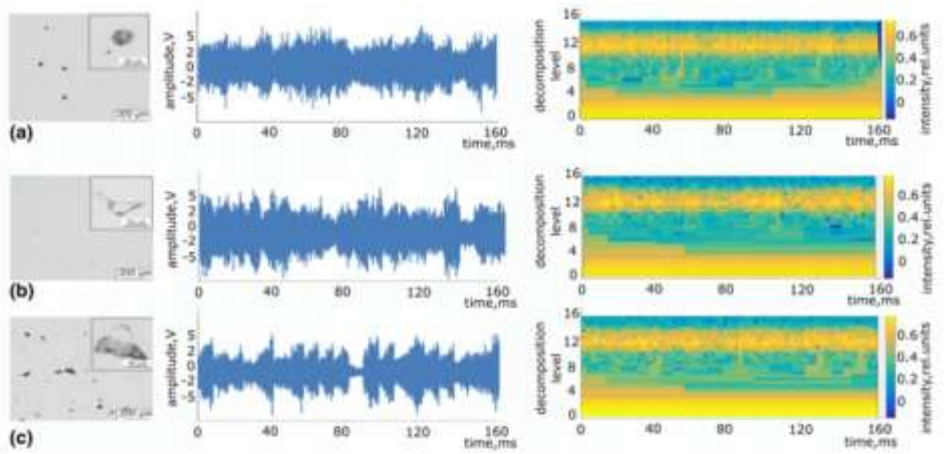


Figure 10 Concept of Convolutional Neuron Network (CNN) [66]

A typical example of current research on classifying signals is in fault diagnosis. Various studies are being to recognize an unusual sound when failure has occurred because it is possible to diagnose the failure simply by installing a sensor such as a microphone outside device [67–73].

In addition, there are many attempts to detect defects during or just after the manufacturing process by classifying sound from that process [74–77]. Similarly, attempts to detect defects in device, such as leaking pipes, are analyzing acoustic emissions via CNN [78].



**Table 2** Test results (in %) for different classes (in rows) versus ground truth (in columns)

Test classes	Ground truth		
	High quality	Medium quality	Poor quality
High quality ( $0.07 \pm 0.02\%$ , 500 mm/s, 79 mm <sup>3</sup> )	<b>74</b>	12	14
Medium quality ( $0.3 \pm 0.18\%$ , 300 mm/s, 132 mm <sup>3</sup> )	12	<b>79</b>	9
Poor quality ( $1.42 \pm 0.85\%$ , 800 mm/s, 50 mm <sup>3</sup> )	11	7	<b>82</b>

Bold values indicate the classification accuracy of the numbers in the diagonal cells

Figure 11 Prediction of the quality of AM products with operation acoustic emission during process [75]

Other current research includes categorizing noise in cities [79], monitoring the condition of structure [80], classifying human activities [81], and identifying genres of music [82].

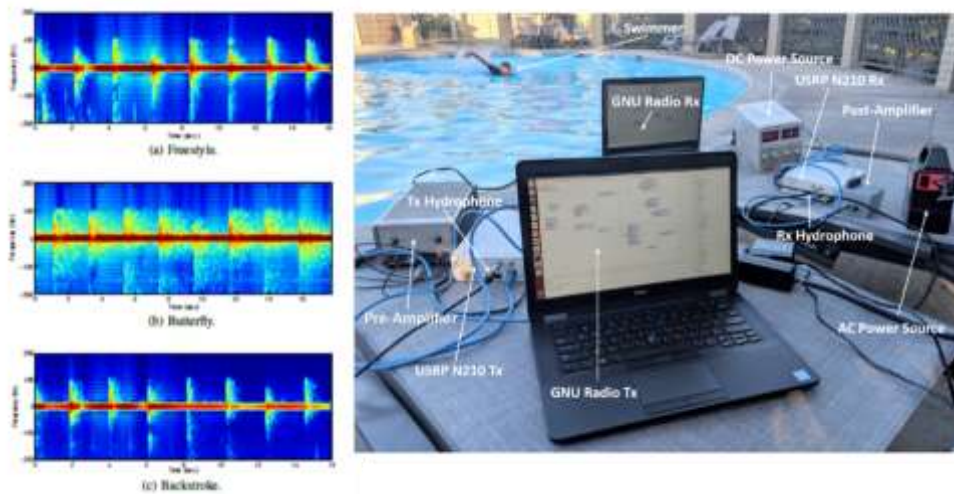


Figure 12 Classifying the human activities with sonar and CNN [81]

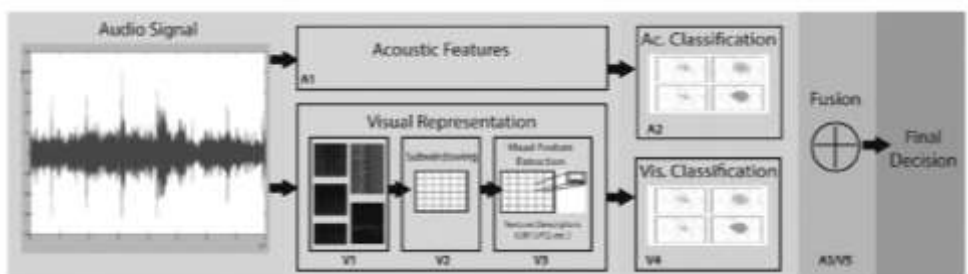


Figure 13 Classifying the genre of music with CNN [82]

In sum, various studies have recently been conducted using CNNs, and these have good results analyzing various sounds. As mentioned previously, a number of studies have been conducted to determine the condition of device using sound; however, no attempt has been made to monitor the condition of multi devices in real time. Therefore, this research aims to develop a sound-based system that can monitor the status of various device in operation simultaneously in a manufacturing process.

## Chap 2. System Modeling

### 2.1 Concept of Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) is an image processing method first proposed by LeCun [66]. CNN has shown significant success in processing images and other forms of data. The convolutional layer of a CNN contains a large number of filters and extracts the characteristics of the input data through these filters. Then local characteristics are extracted by a pooling layer. In the present study, raw data were processed with a Fourier transform to make them 2D; these were the input data for CNN. Another CNN layer, the context one, also used a 2D filter. Note that the input data were from one audio channel, unlike, for instance, an image formed of red, green, and blue channels.

Input data are passed through a convolutional filter to extract the characteristics of each piece of data. The 2D convolutional filter is calculated by the following formula:

$$Y_{i+1} = (Y_i \times F) + b = \sum_N^{-N} \sum_M^{-M} (Y_i \times F) + b \dots\dots\dots(1)$$

where  $Y_i$  and  $Y_{i+1}$  are the data before and after passing through the filter, respectively;  $F$  is the filter; and  $b$  is the bias. The set of  $Y_{i+1}$  is called the feature map.

A pooling layer, which extracts important local information from the feature map, is typically applied after the convolutional layer. However, as it passes through the pooling layer, the dimensions of the feature map are reduced. Average or max pooling layers are commonly used; in the present study, we used the latter. The max pooling operation extracts only the maximum size of the filter kernel in the feature map. The geometry extracted from the filter kernel is obtained as follows:

$$A = [a_{ij}] \quad (i, j \leq n) \dots\dots\dots(2)$$

$$\text{maxpooling}(A) = \max(a_{ij}) \dots\dots\dots(3)$$

where  $A$  is the filter kernel and  $a_{ij}$  is an element of the filter kernel. A completely connected layer and softmax classification were added to classify the data by alternately using convolution and pooling layers. A typical 2D CNN structure is discussed below.

### 2.1.1 Dropout

Dropout is a technology that can reduce data overfitting. Particularly when training a small neural network, dropout can prevent a reduction in performance, providing an easy and effective way to solve this problem. In the present study, dropout techniques were applied during training to prevent hollowing out, with repeated extraction of the same function. Some hidden neurons were set to zero so that they were not included in feedforward learning.

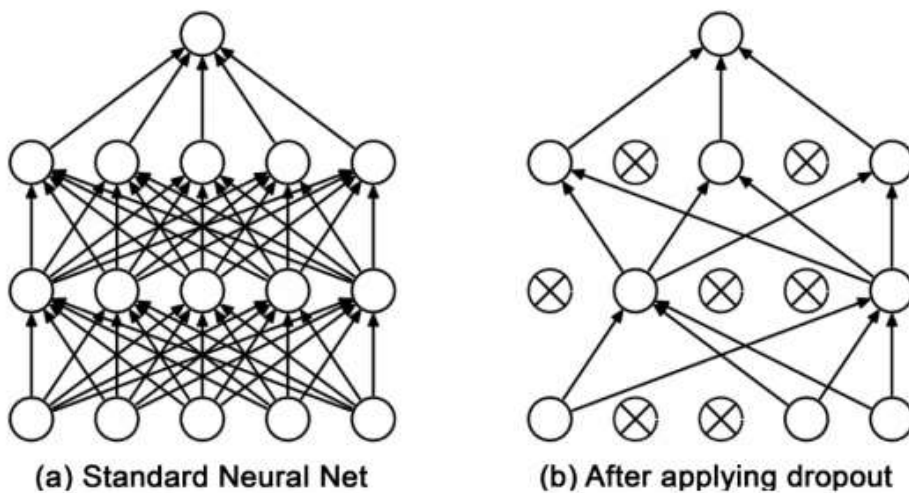


Figure 14 Concept of dropout [83]

### 2.1.2 Softmax classifier

Softmax regression is typically implemented as the top layer of the neural network for multi-state classification. Information derived from multiple hidden layers is used as input for supervised classifiers according to global back-propagation optimization. In the present study, we used softmax regression as a mechanical health status classifier in the network. Training samples are represented by  $x(i)$  and their label set is  $y(i)$  where  $i = 1, 2, \dots, K$  is the number of training samples.

$$x(i) \in R^{N \times L}, y(i) \in \{1, 2, 3, 4, \dots, K\} \dots\dots\dots(4)$$

(K is the number of categories labeled)

For  $x(i)$ , input sample, softmax regression can estimate the probability as

$$P(y(i) = j \mid x(i)) \text{ for each label } j \ (j = 1, 2, 3, \dots, K) \dots\dots\dots(5)$$



The estimated probability of  $x(i)$  belonging to each label can be obtained according to the hypothesis function,

$$\begin{aligned} \text{Softmax}(x(i)) &= [p(y(i)=1|x(i)), p(y(i)=2|x(i)), \dots, p(y(i)=K|x(i))] \\ &= \left[ \frac{e^{x(1)}}{\sum_{j=1}^K e^{x(j)}}, \frac{e^{x(2)}}{\sum_{j=1}^K e^{x(j)}}, \dots, \frac{e^{x(i)}}{\sum_{j=1}^K e^{x(j)}} \right] \dots\dots\dots(6) \end{aligned}$$

This classifier verifies that the output is positive and the sum is 1, so that the output of the network can be interpreted as the probability of each class.

## 2.2 Fourier transform

In mathematics, a Fourier transform (FT) is a mathematical transform that decomposes a function (often a function of time, or a signal) into its constituent frequencies. The Fourier transform of a function of time is a complex-valued function of frequency, whose magnitude (absolute value) represents the amount of that frequency present in the original function, and whose argument is the phase offset of the basic sinusoid in that frequency.

The Short-time Fourier transform (STFT), is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.

In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. This reveals the Fourier spectrum on each shorter segment. One then usually plots the changing spectra as a function of time, known as a spectrogram or waterfall plot.

With this STFT, arbitrary 1-D digital data with discrete value can be transformed from 1-D time domain to 2-D frequency domain. And it is usually used to analyze the time domain signals such as sound.

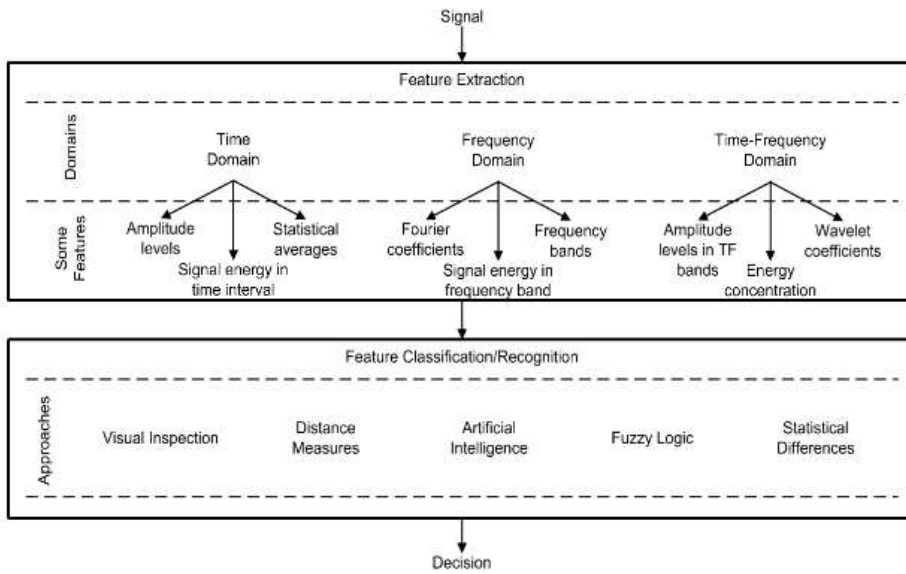


Figure 15 Signal processing method for pattern classification [84]

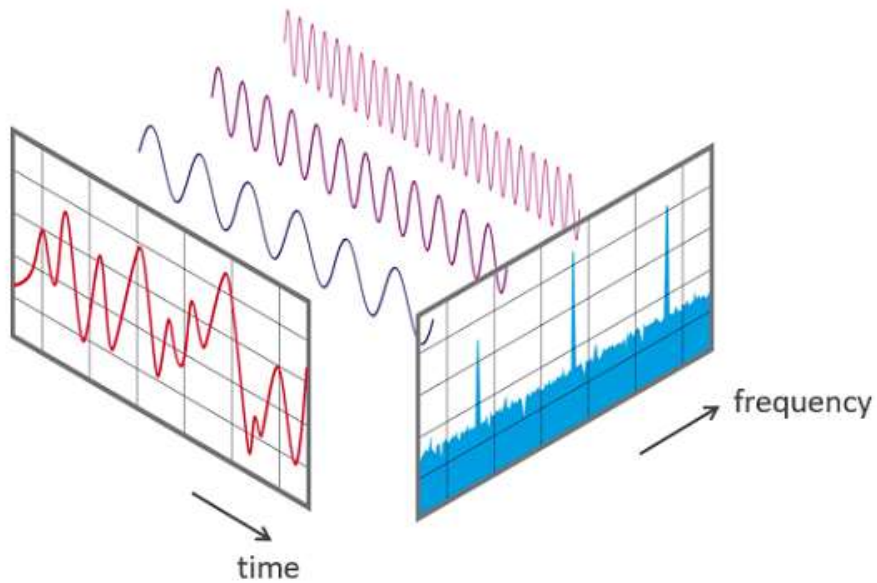


Figure 16 Concept of Fourier Transform [85]

### 2.2.1. Continuous-time STFT

Simply, in the continuous-time case, the function to be transformed is multiplied by a window function which is nonzero for only a short period of time. The Fourier transform (a one-dimensional function) of the resulting signal is taken as the window is slid along the time axis, resulting in a two-dimensional representation of the signal. Mathematically, this is written as:

$$\text{STFT}\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-i\omega t} dt \dots\dots\dots(7)$$

where  $w(t)$  is the window function, commonly a Hann window or Gaussian window centered around zero, and  $x(t)$  is the signal to be transformed (note the difference between the window function  $w(t)$  and the signal  $x(t)$ ). The STFT is essentially the Fourier transform of a complex function representing the phase and magnitude of the signal over time and frequency. Often phase unwrapping is employed along either or both the time axis, to suppress any jump discontinuity of the phase result of the STFT. The time is normally considered to be "slow" time and usually not expressed in as high resolution as time

### 2.2.2. Discrete-time STFT

In the discrete time case, the data to be transformed could be broken up into chunks or frames (which usually overlap each other, to reduce artifacts at the boundary). Each chunk is Fourier transformed, and the complex result is added to a matrix, which records magnitude and phase for each point in time and frequency. This can be expressed as:

$$\text{STFT}\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n - m]e^{-j\omega n} \dots\dots\dots(8)$$

likewise, with signal  $x[n]$  and window  $w[n]$ . In this case,  $m$  is discrete and  $\omega$  is continuous, but in most typical applications the STFT is performed on a computer using the fast Fourier transform, so both variables are discrete and quantized.

$$\text{spectrogram}\{x(t)\} \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n - m]e^{-j\omega n} \dots\dots\dots(9)$$

The magnitude squared of the STFT yields the spectrogram representation of the Power Spectral Density of the function:

## 2.3 Log-mel spectrogram

Log-mel spectrogram is one of the most popular spectrogram based on STFT. Studies have shown that humans do not perceive frequencies on a linear scale. We are better at detecting differences in lower frequencies than higher frequencies. Therefore, Mel spectrum is more suitable in human's auditory sense characteristic that presents the linear distribution under the 8000 Hz and the logarithm growth above the 8000 Hz, we utilize this point to obtain the Log-Mel spectrum static. The relationship between the Mel spectrum and the frequency is shown as

$$f_{mel} = 2595 \times \log_{10} \left( \frac{1+f}{700} \right) \dots\dots\dots(10)$$

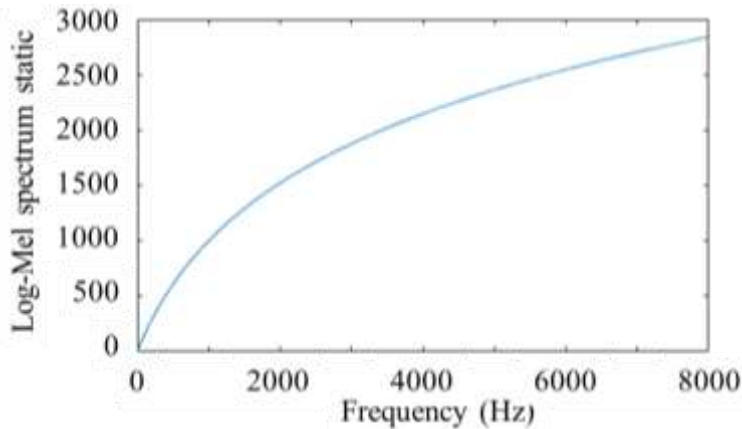


Figure 17 Relation between frequency and log-mel spectrum static

We adopt the number of 40 filterbanks to process the raw signal under the control of the 16 kHz sample rate and the length of FFT is set to 512.

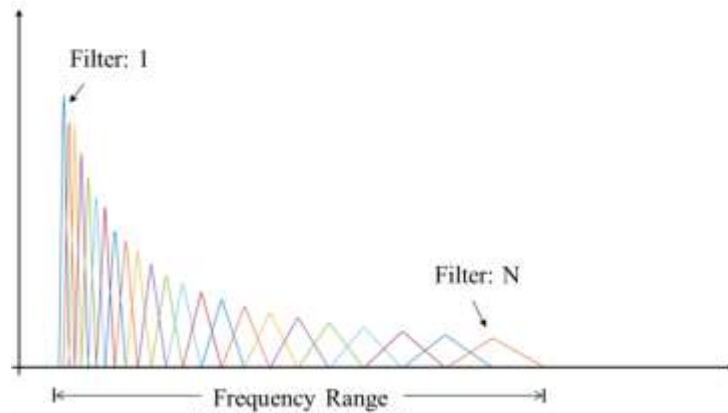


Figure 18 Filterbank for log-mel spectrogram

Furthermore, we choose the hamming window which is taken the window length of 25 ms and the window shift of 10ms to add into the signal. Before gaining the 40 Mel-filterbank vectors, we also select the lower frequency of 50 and the upper frequency of 7000. Then we will take the signal to feed into the filterbanks to get the  $H_m(k)$ , which is shown as

$$\begin{aligned}
H_m(k) = & \\
& \frac{k - f(m - 1)}{f(m) - f(m - 1)} \text{ (if } f(m - 1) \leq k \leq f(m) \text{) or} \\
& \frac{f(m + 1) - k}{f(m + 1) - f(m)} \text{ (if } f(m) \leq k \leq f(m + 1) \text{) or} \\
& 0 \text{ (if } k < f(m - 1) \text{ or } f(m + 1) < k \text{) .....(12)
\end{aligned}$$

According to the results of computing, we will get the outputs from the filterbanks, and then multiply the energy spectrum is used by the STFT processed from the raw signal, which is shown as

$$\log - \text{melspec}(m) = \sum_{k=f(m-1)}^{f(m+1)} \log(H_m(k) \times |X(k)|^2) \text{ .....(13)}$$

where the  $|X(k)|^2$  describes the energy spectrum in the points of  $k$ th energy,  $m$  is the number of the filterbanks and  $k$  is the point of the FFTs.



## 2.4 Proposed architecture

Deep neural networks are capable of the adaptive capture of information related to facial expressions from raw input signals through multiple nonlinear transformations and approximate, complex, nonlinear functions; such networks are typically used as the main CNN architecture. To this architectural base, algorithms can be added to efficiently train networks and improve diagnostic performance. Figure 19 shows the structure of the proposed network for acoustic monitoring. In the proposed framework, raw collected data are converted into 2D form and used as the model input, and no prior expertise in signal processing and fault diagnosis required.

A zero-adjustment operation is implemented to ensure that the geometry map dimensions are not changed. Pooling layers are usually used in deep networks to reduce the number of parameters and accelerate the training process while retaining important features of the information. Pooling layer decisions depend on specific fault diagnosis problems and their data sets. In most cases, the average pooling layer is used between two remaining building blocks. Finally, the learned features extracted by the system are passed to fully connected layers and softmax regression to estimate the failure categories. Batch normalization can accelerate the training process, in particular for deep learning, and has demonstrated good

performance in recent studies. In the present study, batch normalization was used after each convolutional layer. In addition, we used the rectified linear unit activation function in the network. Because it does not suffer from gradient diffusion during the training process, better performance can typically be achieved, in particular in a deep structure.

Cross entropy function is used as loss function in the learning process. Back-propagation (BP) algorithms are applied to all weight updates in the layers and used the stochastic gradient drop optimization method during training. When a lot of training data are required, useful training samples can be generated by data enhancement. Multiple CNN building blocks can potentially be stacked in the network to ensure better functional extraction through a deeper structure.

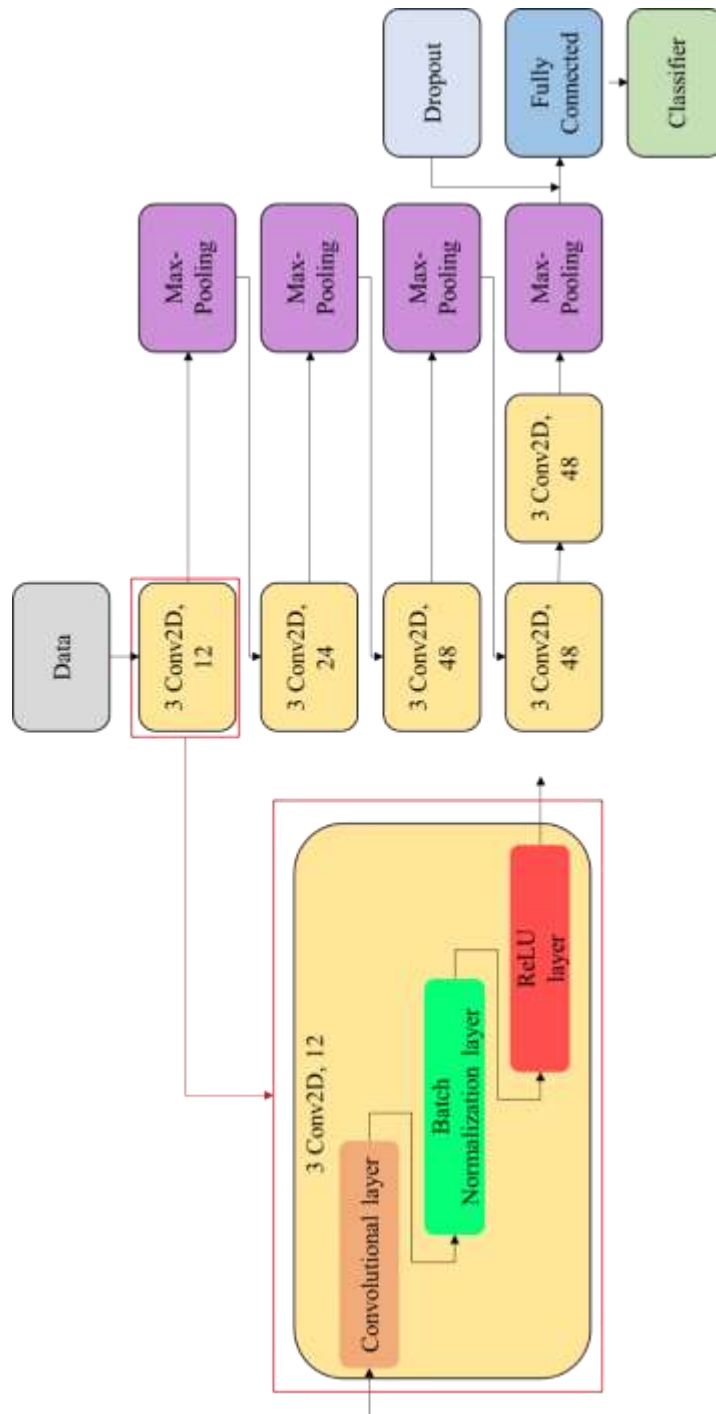


Figure 19 Proposed convolutional neural network architecture for operational monitoring.

## **2.5 Concept of monitoring system**

The algorithms proposed in the present study are shown in Figure 20. First a raw signal is acquired to form a data set for training. The raw signal may be obtained directly from a mic array in advance, but it is also possible to use a virtual data set, which will be described in Chapter 5. Data were trimmed to a duration of 1 sec and were labeled to allow data sorting. Data augmentation process was conducted, including translating and adjusting the time scale, and added background noise, giving the final data set.

Then data composed in this way were converted into 2D form with a log-mel spectrogram. Log-mel spectrogram is a kind of wavelet transform based on short-time Fourier transform (STFT) that allocates a frequency band area to a human audible frequency and converts its size to a logarithmic scale; it can provide results similar to what a person hears. Thus, it is mainly used in the sound recognition field and achieves very high performance, in particular in classification [86]. Data set was sorted according to the operating status of each devices.

To use trained CNN, a machine's operating sound is recorded through a mic array and, as in the process of preparing the training data set, trimmed to samples 1 sec long and converted into 2D data with the log-mel spectrogram. The CNN then calculates the predicted

probability of each operational status of each device, thus predicting the most likely operational status of the device.

In our proposed monitoring system, each device has its own CNN that classifies the operational status of that item, and the system monitors all device in parallel. Each CNN classifies only the operational status of its one item, regarding the sound of all other items as noise. The advantage of this monitoring system is that, even if several devices are operating simultaneously, it is possible to identify a piece of device and its status provided only that operating sound is detected and to monitor unlimited devices simultaneously.

The specifications of the hardware and software used in this research are listed in Table 1

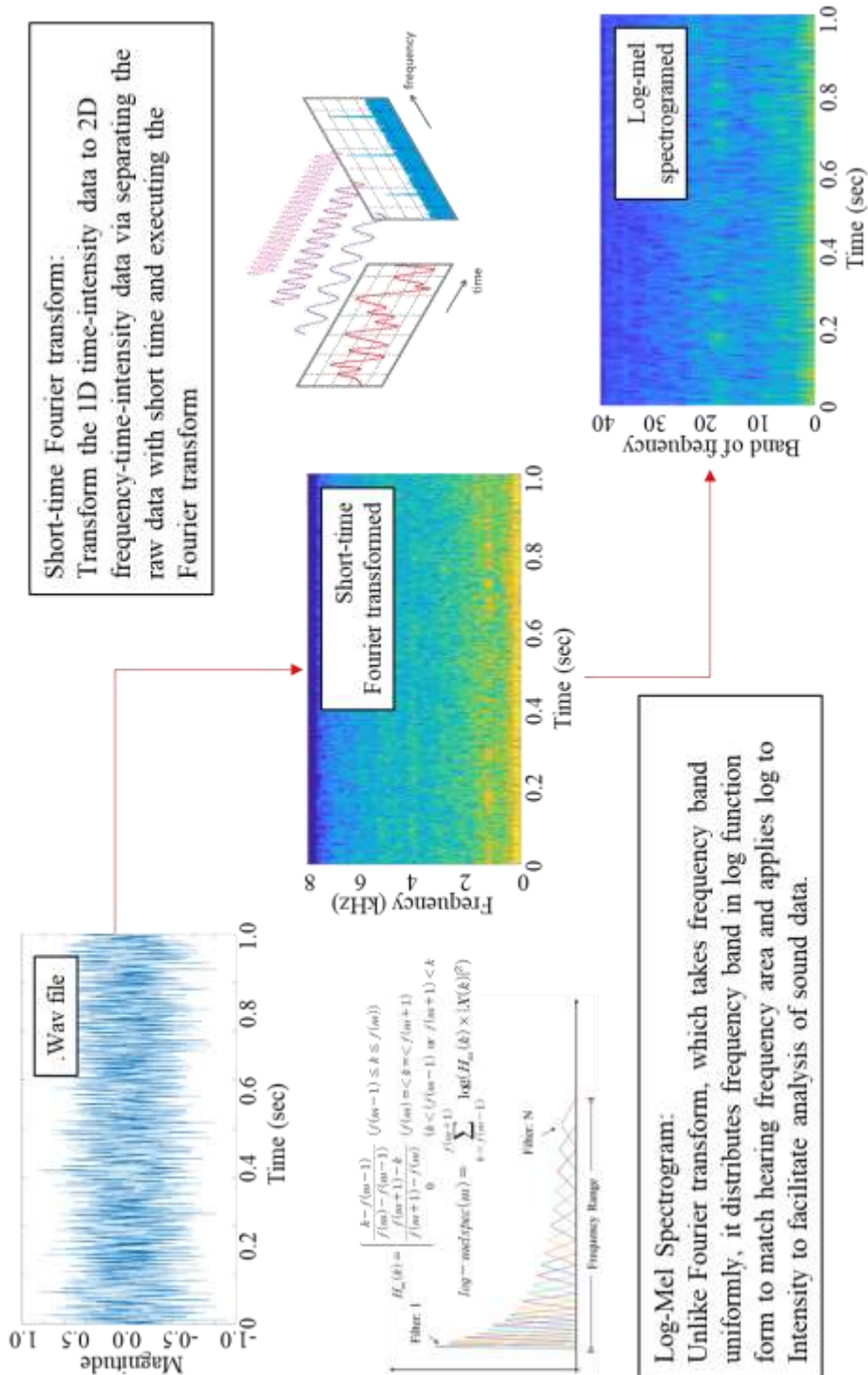


Figure 20 Concept diagram of sound data conversion from 1D to 2D

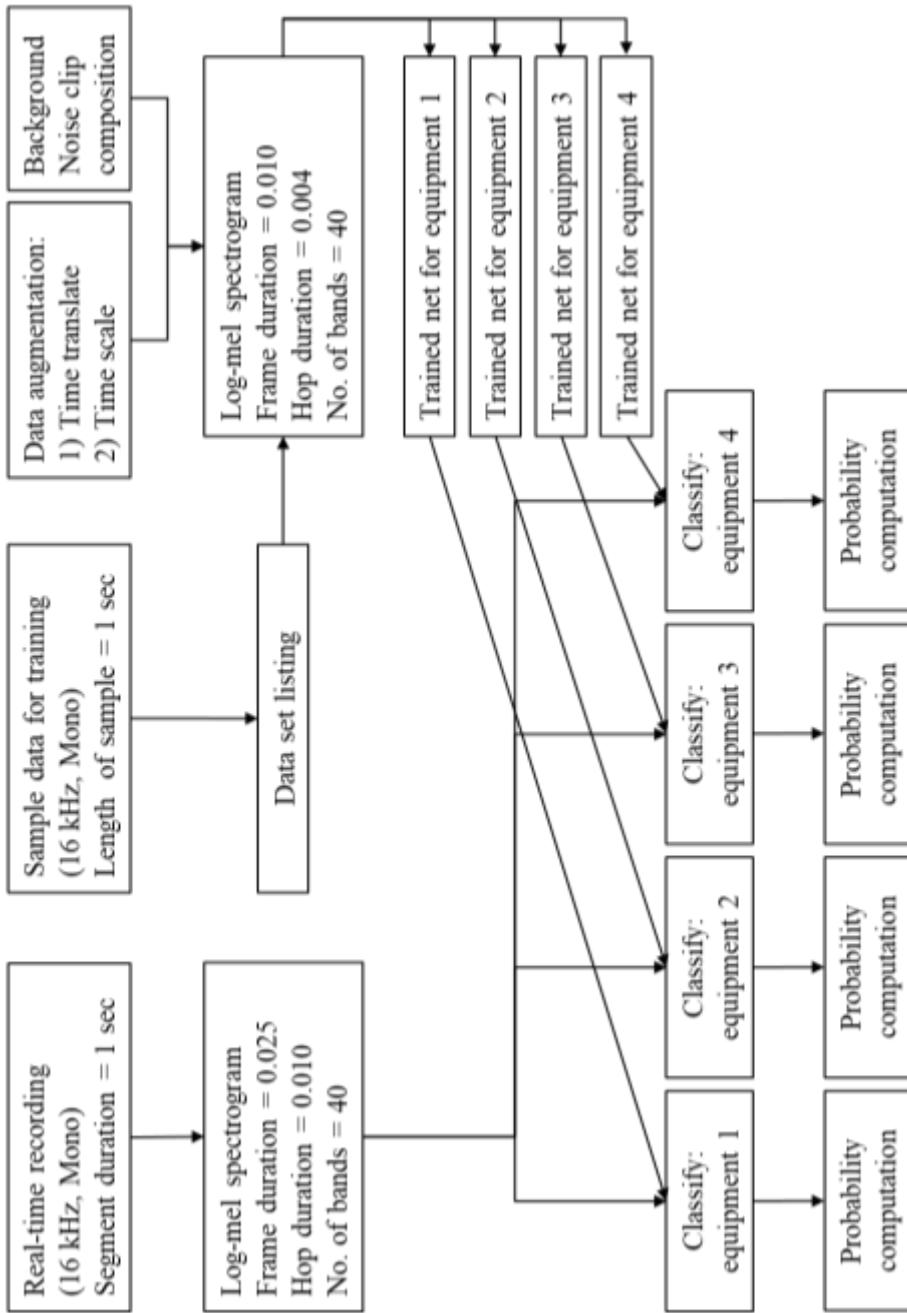


Figure 21 Schematic of the monitoring system

Table 1 Specification of hardware/software

Specification	Value
Operating System	Microsoft® Windows® 10 Home
Software Platform	MATLAB R2020a
System RAM	Samsung® 32 GB (DDR3)
Processor Type (CPU)	Intel® core i7-4790 (3.9 GHz)
Graphics Card (GPU)	NVIDIA® GeForce GTX 750  (RAM: 1GB)



## **2.6 Parallel and independent system**

Sound has the characteristics of being recorded differently by the location of listening and the setting of the microphone, but the human ear is all recognized by the same kind of sound. Also, sound can generally be propagated by bypassing obstacles, so when recording sound, various noise is inevitably mixed. This is important in the process of building a system that recognizes real-world sound system. Moreover, sounds generated when the equipment is operated tend to repeat slightly different sounds while being similar. Given all these considerations, it was judged that it was not efficient to separate/recognize these sounds by creating a single large neural network.

In this situation, we considered how real people perceive sound. In the end, people usually listen to a variety of sounds at the same time, but when analyzing them specifically, they tend to focus on one sound. This is because in all cases the same model can be used if individual sounds are classified as sounds of a particular equipment and other noise.

For this reason, the system was implemented by building and merging multiple simple independent artificial neural networks, which are solely responsible for individual equipment, using the characteristics

of overlapping sounds that retain their own characteristics when multiple sounds were added.

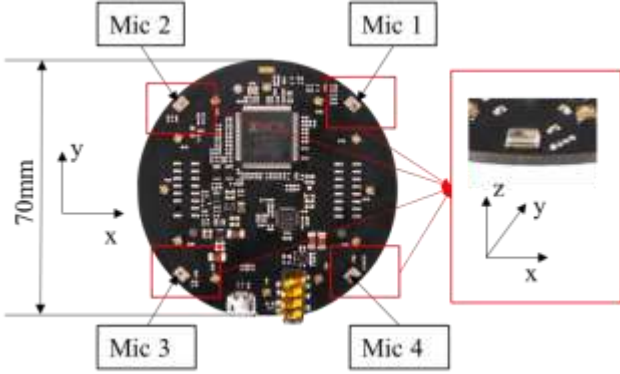
## **Chap 3. Development of the monitoring system**

### **3.1 Hardware of monitoring system**

In this chapter, we describe the creation of a real monitoring system and explain how we obtained the data required. We explain the type of information obtained through this recording process and describe the results from analyzing it through a given algorithm.

The process begins with recording sound from a workspace. The detailed specification of the recording device (Respeaker Mic Array 2.0) is given in Table 2. Information of target devices is as shown in Table 3 and example of the recorded signal is shown in Figure 22. The number of outputs from the mic array is five: Four signals are the raw signals from each mic, and the fifth is output by the digital signal processor. Omnidirectional mics are used, so a phase difference exists that depends on the time differences in arrival and the received signals are almost identical.

Table 2 Specification of mic array

Specification	Value
Name	Respeaker Mic Array 2.0
No. of Mic	4 (Output: 5ch)
Sensitivity	26 dBFS (Omnidirectional)
Diameter	$\Phi 70$ mm
Max sample rate	48 kHz
Digital signal processor	XMOS XVF-3000
Recording Program	Audacity 2.3.3
Appearance of mic array	
 <p>The diagram illustrates the physical appearance of the Respeaker Mic Array 2.0. It is a circular black PCB with a diameter of 70mm. Four microphones are positioned at the corners of a square on the board, labeled Mic 1, Mic 2, Mic 3, and Mic 4. A coordinate system is defined with the x-axis pointing right, the y-axis pointing up, and the z-axis pointing out of the page. An inset image shows the physical device with a red dot indicating a specific location on the board.</p>	

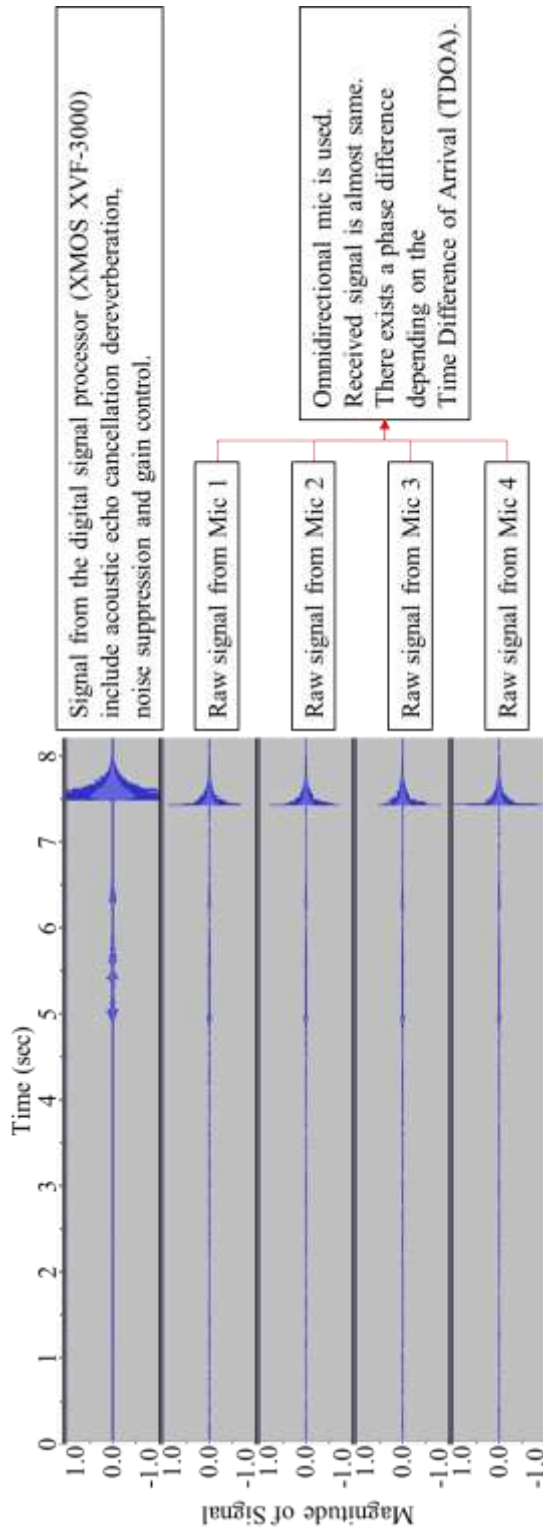


Figure 22 Signals received and processed from the by mic array

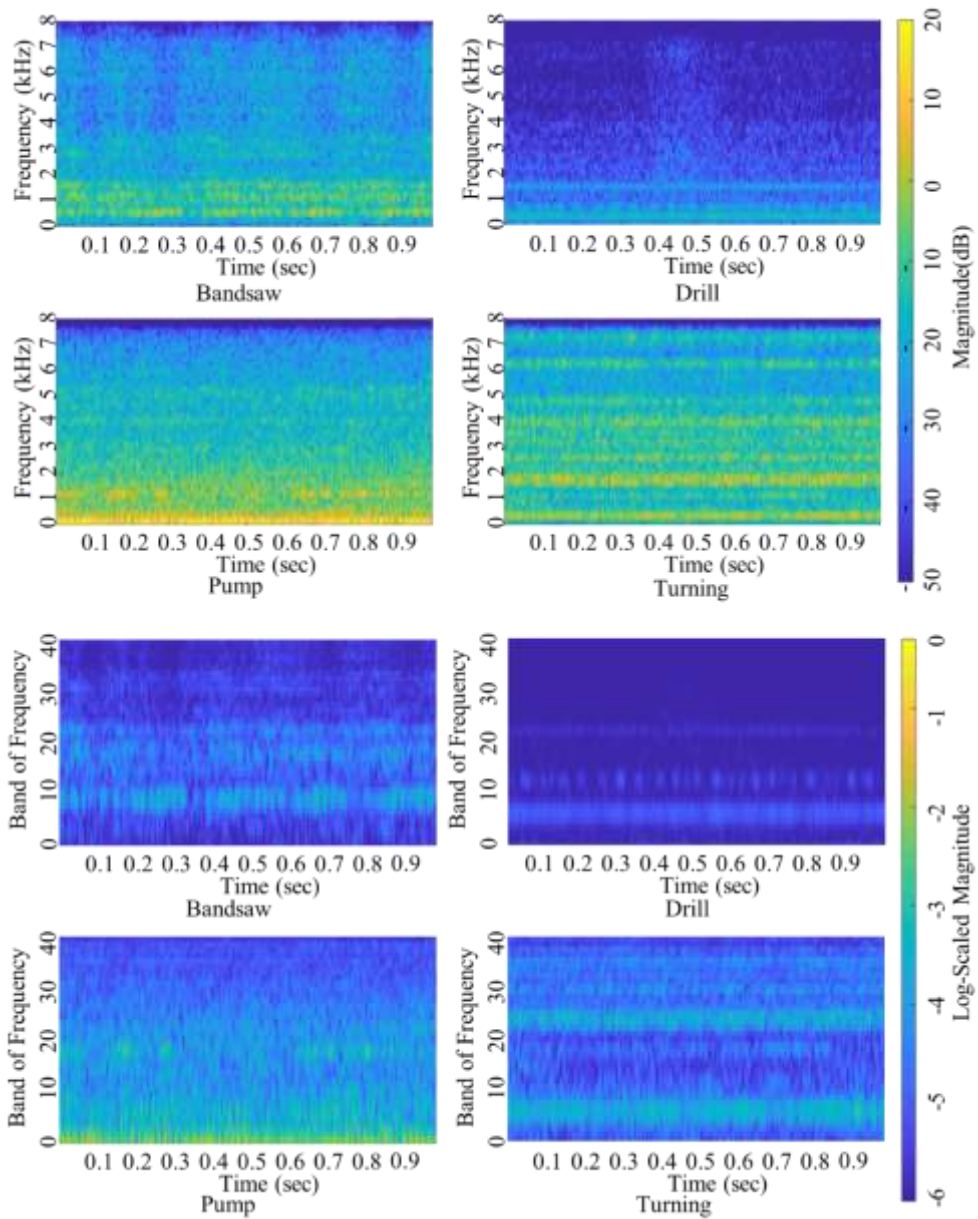


Figure 23 Operating sounds from the target devices after processing by STFT and log-mel spectrogram.

Table 3 Monitoring target device

Name	Intensity of sound	
Bandsaw	80 dBA	
Drill	65 dBA	
Pump	87 dBA	
Turning	95 dBA	

※ Intensity of background sound: 55 dBA

### **3.2 Training with actual data**

A bandsaw, a drill, a pump, and a turning were selected as targets for monitoring. These devices were placed in the laboratory in arbitrary positions. Figure 23 shows a recording results of their operating sounds post-processed with STFT and log-mel spectrogram.

To make the data set for training, recording was performed according to the scheme in Figure 24 and Table 4. Two mic arrays were installed, and two signals were collected from each: the raw signal from mic 1 and the digital signal processor output. Recording lasted 800 sec and recorded sound whose total length is 3200sec was divided the results into 3200 files of 1 sec duration each. To ensure variety in the data, the mic arrays were installed at arbitrary locations.



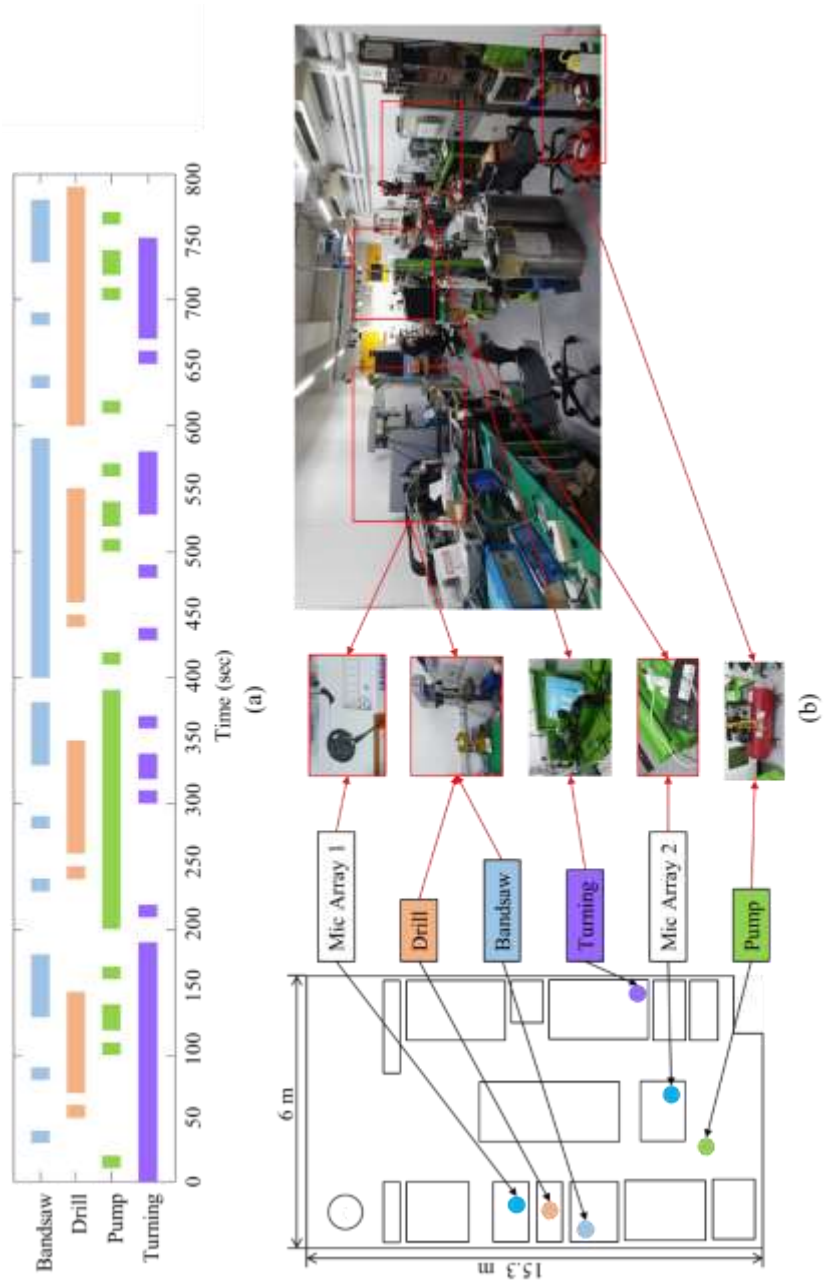


Figure 24 Experimental setup for acquiring sample data to train the monitoring system.

(a) The times at which the devices were operated.

(b) The position of each device.

Table 4 Specification of dataset for data training

Specification	Value
Total number of data (recording time)	3200 files (800 sec)
Data format	.wav file with 1 sec length
Number of target device	4 (bandsaw, drill, pump, turning)
Number of channel	4 (2ch per each mic array and 2 mic arrays)

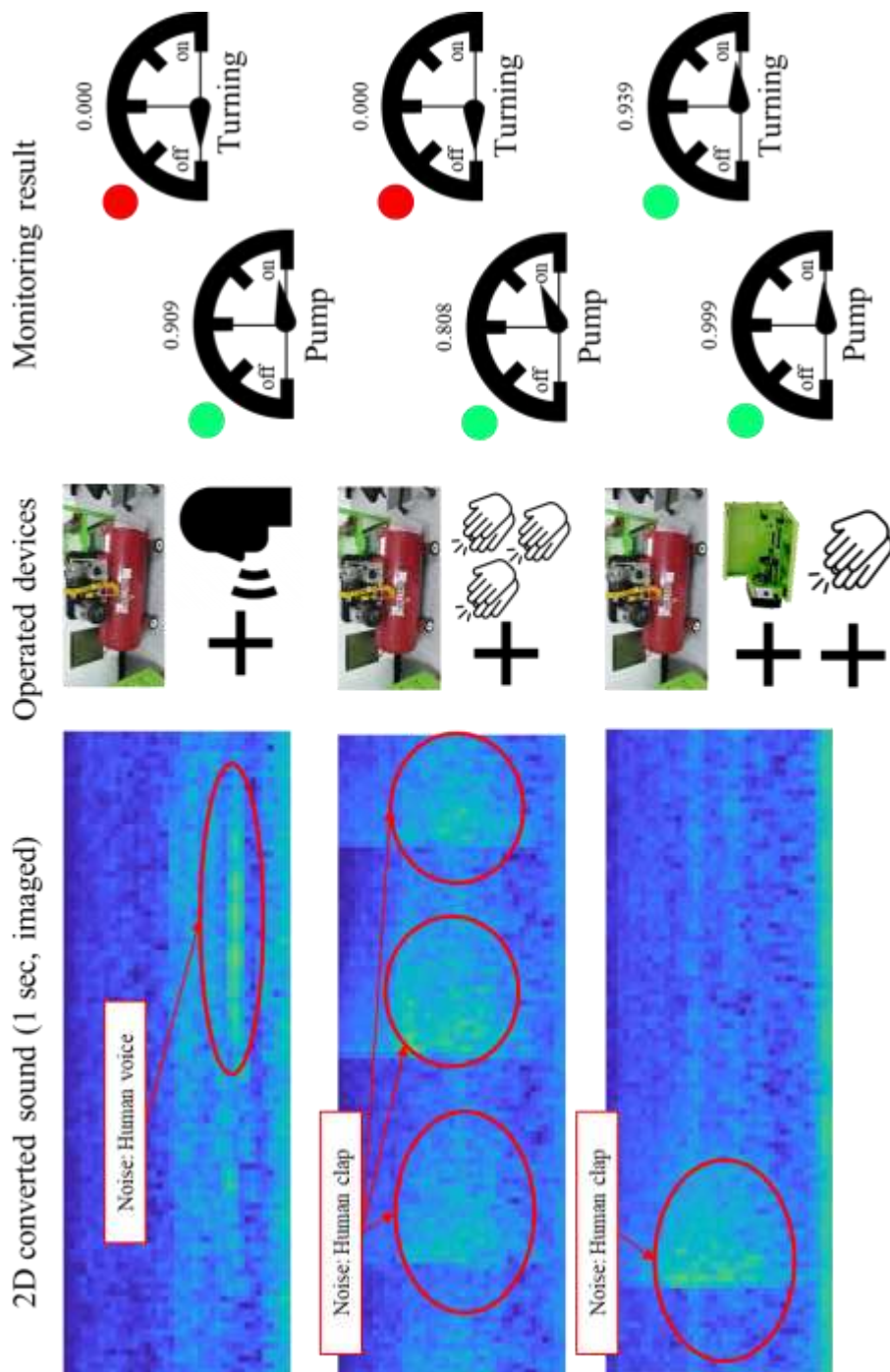


Figure 25 Results of operation monitoring

Based on that recorded sound, monitoring system was trained. Figure 25 shows the results of using the system to monitor the pump and turning: The system was aware of the sound of the device and could detect that it was operating without problems when two devices were operating simultaneously or when the devices were operating with noise (such as a person clapping).

### **3.3 Performance evaluation of monitoring system**

To evaluate the performance of the system, test data set was created according to the scheme in Figure 26 and Table 5. As before, two mic arrays were installed, and two signals from each (the raw signal from mic 1 and the digital signal processor output) were recorded for 160 sec, giving a total of 640 sec of recorded sound. The data were divided into 640 .wav files of 1 sec duration each. To ensure variety in the data, the mic arrays were installed in new, arbitrary locations. Performance of the monitoring system was evaluated by comparing the results for this test set to the operational scheme; performance was defined as the proportion of predictions that matched the operational status of each device.

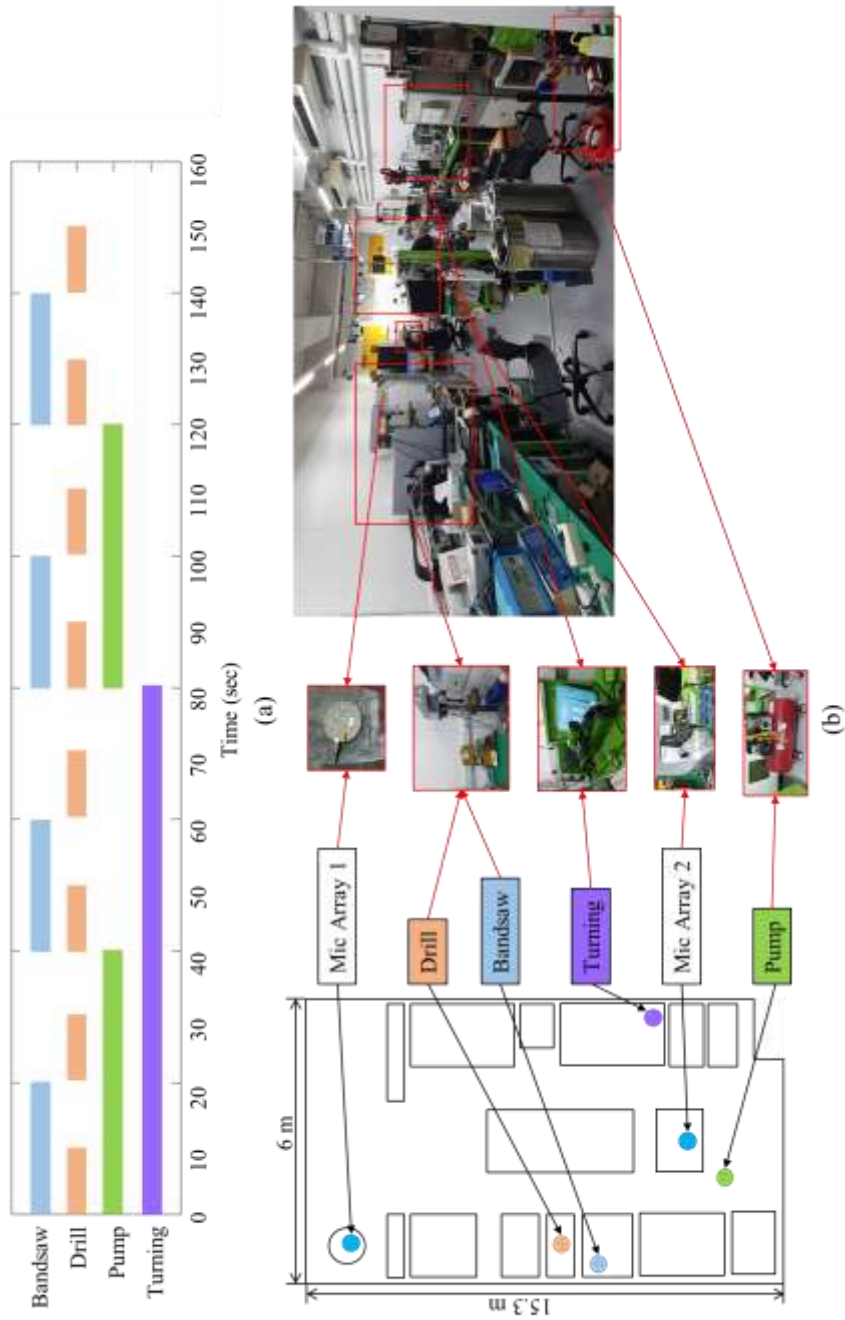


Figure 26 Experimental setup for acquiring the test data set for evaluating the monitoring system. (a) The times at which the devices were operated. (b) The position of each device.

Table 5 Specification of standard data set

Specification	Value
Total number of data	640 files (160 sec)
Data format	.wav file with 1 sec length
Number of target device	4 (bandsaw, drill, pump, turning)
Number of channel	4 (2ch per each mic array and 2 mic arrays)

The performance results are shown in Figure 27. The accuracy of the monitoring system was approximately 70%, 53%, 95%, and 91% for the bandsaw, drill, pump and turning, respectively. This shows that the system could recognize the sounds of the bandsaw, pump, and turning well but could not recognize the sound of the drill. This is because the drill was very quiet, so the CNN did not properly extract its characteristics during the training process; the sound of the drill was disguised by other device sounds even when it was operating.



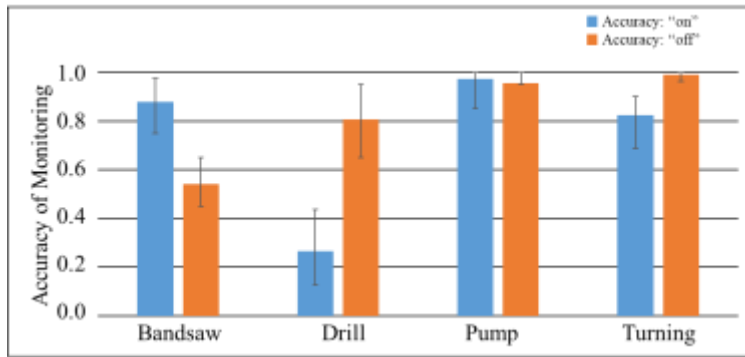


Figure 27 Performance of the monitoring system based on real data  
 (Training data set consisted of 3,200 samples over a target  
 frequency range of 50-7,000 Hz)

### **3.3.1 Recognition performance at the different mic position**

The intensity of sound decreases in proportion to the square of the distance. This shows that the location of the mic and the resulting distance change from the music source are closely related to the strength of the signal input through the mic. This means that when the various sounds are combined, location of mic can be important variable that can directly affect to recognition performance.

To deduce the relation between location and distance from sound source (device), monitoring was executed with 6 different point and 8 different operation condition for 10 sec each, total monitoring time was 480 sec. Detail condition is as Figure 28 and the results are as shown in Table 6.

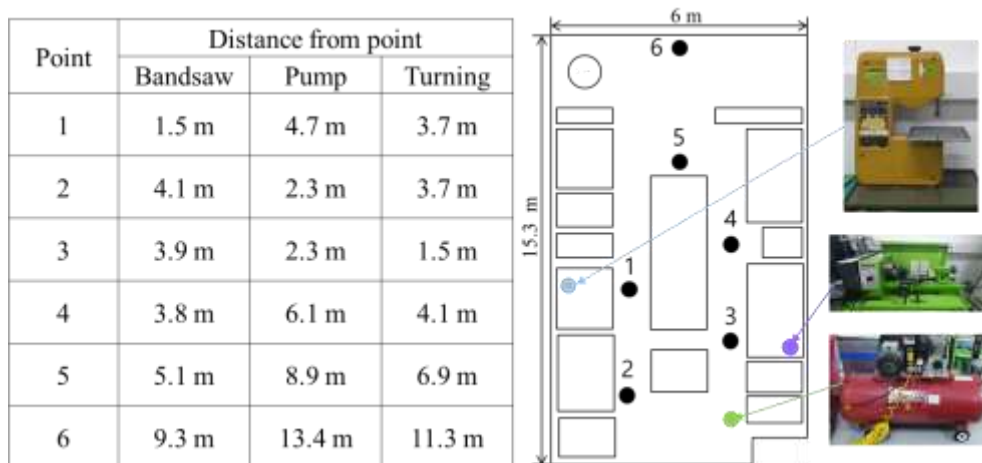


Figure 28 Performance test with different mic position

Table 6 Recognition performance with different mic position

Point	Bandsaw			Pump			Turning		
	overall	on	off	overall	on	off	overall	on	off
1	75.0%	95.0%	55.0%	97.5%	95.0%	100.0%	92.5%	85.0%	100.0%
2	72.5%	92.5%	52.5%	98.8%	97.5%	100.0%	97.5%	95.0%	100.0%
3	71.3%	90.0%	52.5%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
4	63.8%	80.0%	47.5%	98.8%	97.5%	100.0%	100.0%	100.0%	100.0%
5	68.8%	75.0%	62.5%	98.8%	97.5%	100.0%	100.0%	100.0%	100.0%
6	62.5%	70.0%	55.0%	93.8%	87.5%	100.0%	85.0%	70.0%	100.0%

### **3.3.2 Recognition performance with different operation mode**

Device can operate differently depending on its operating mode and environment. It means, device can emit the different operation sound. Existing methods of monitoring responding to specific conditions often fail to function properly due to changes in these signals. However, for monitoring systems based on artificial intelligence and CNN, monitoring is possible even if there are some changes in signal.

To check the recognition performance in these situations, monitoring was executed for bandsaw at the different situation, with 2 operation speed (130 m/min, 200 m/min) and 2 materials (wood, NC nylon). Recognition accuracy was 92% and detail result was same as Figure 29.

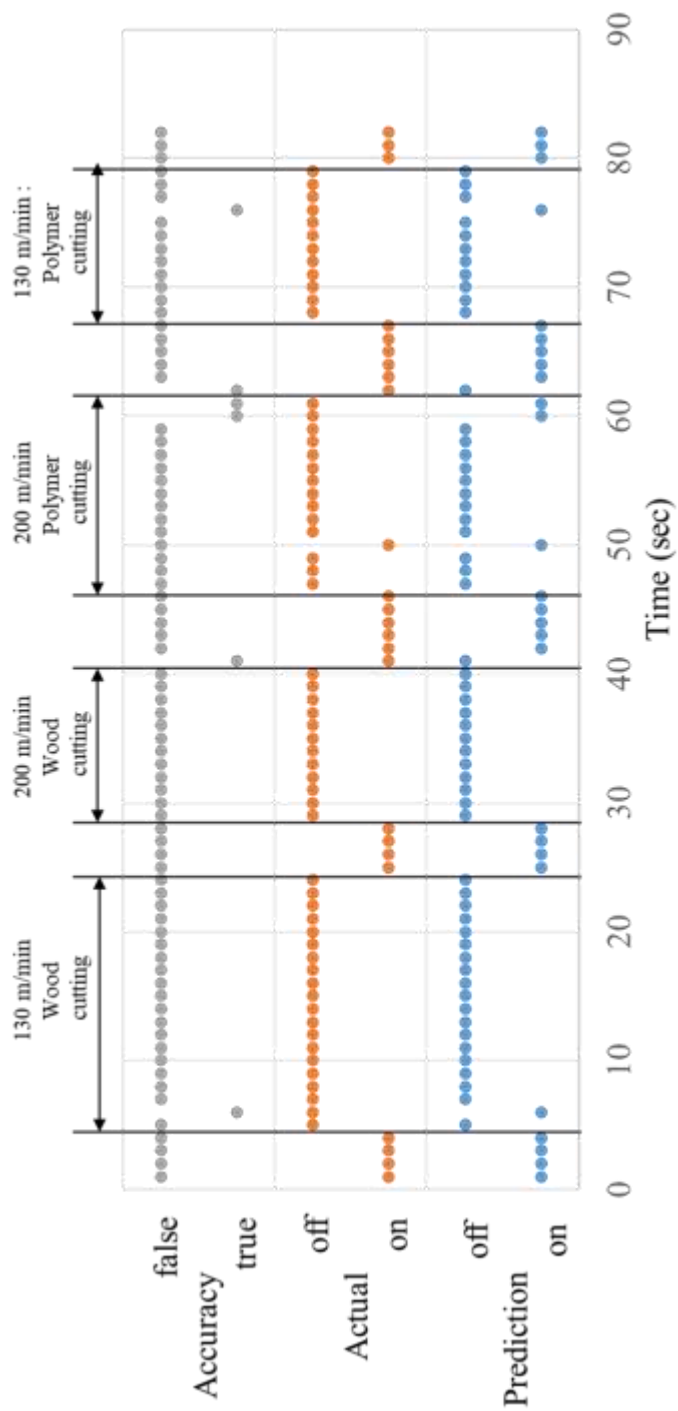


Figure 29 Bandsaw operation monitoring with different operation environment

## **3.4 Enhancement of performance**

### **3.4.1 Modifying target frequency range**

To increase the recognition rate for the device, changing the frequency targeted for monitoring is tried. Figure 30 shows the results after conversion the sound used in the test data set to the frequency domain using STFT. The operating sound of device is usually concentrated at a specific frequency, and the performance of the monitoring system should improve if the system is set to focus on that frequency. In the case of the bandsaw, the target frequency range was changed from 50 - 7,000 Hz to 10 - 1,500 Hz, mainly because of its distinct characteristics in the low-frequency area, below 1,500 Hz. The accuracy of the monitoring system thus improved from 71% to 85%, as can be seen in Figure 31.

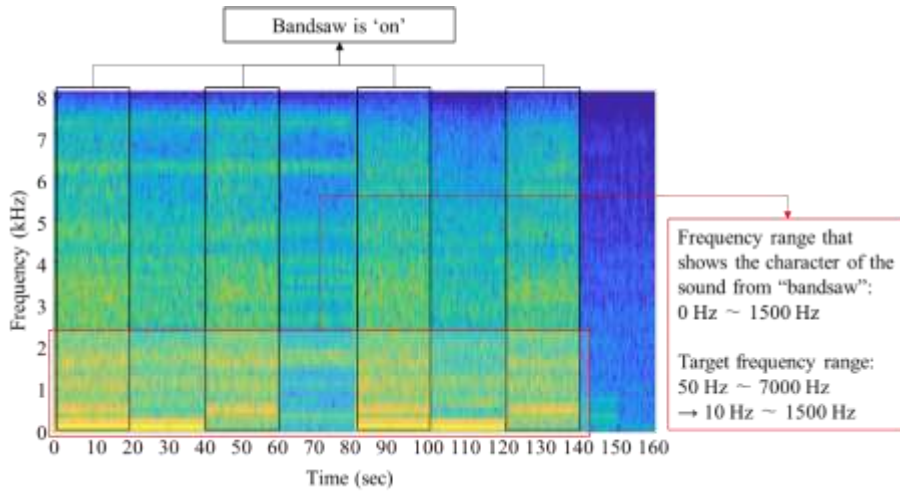


Figure 30 Fourier transformed operating sound

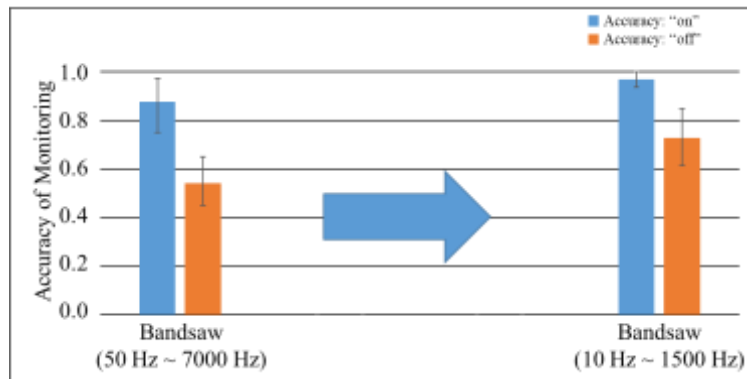


Figure 31 Performance evaluation of monitoring system with different target frequency ranges (left) 50 Hz ~ 7000 Hz and (right) 10 Hz ~ 1500 Hz

As can be seen above, performance can be improved just by changing the frequency range. To see how the performance changes due to the frequency range change are shown, experiments were conducted in more diverse frequency areas, and the results were same as Figure 32. Based on the results of the experiments shown in Figure 33 the values were estimated as three dimensions to see the trend.



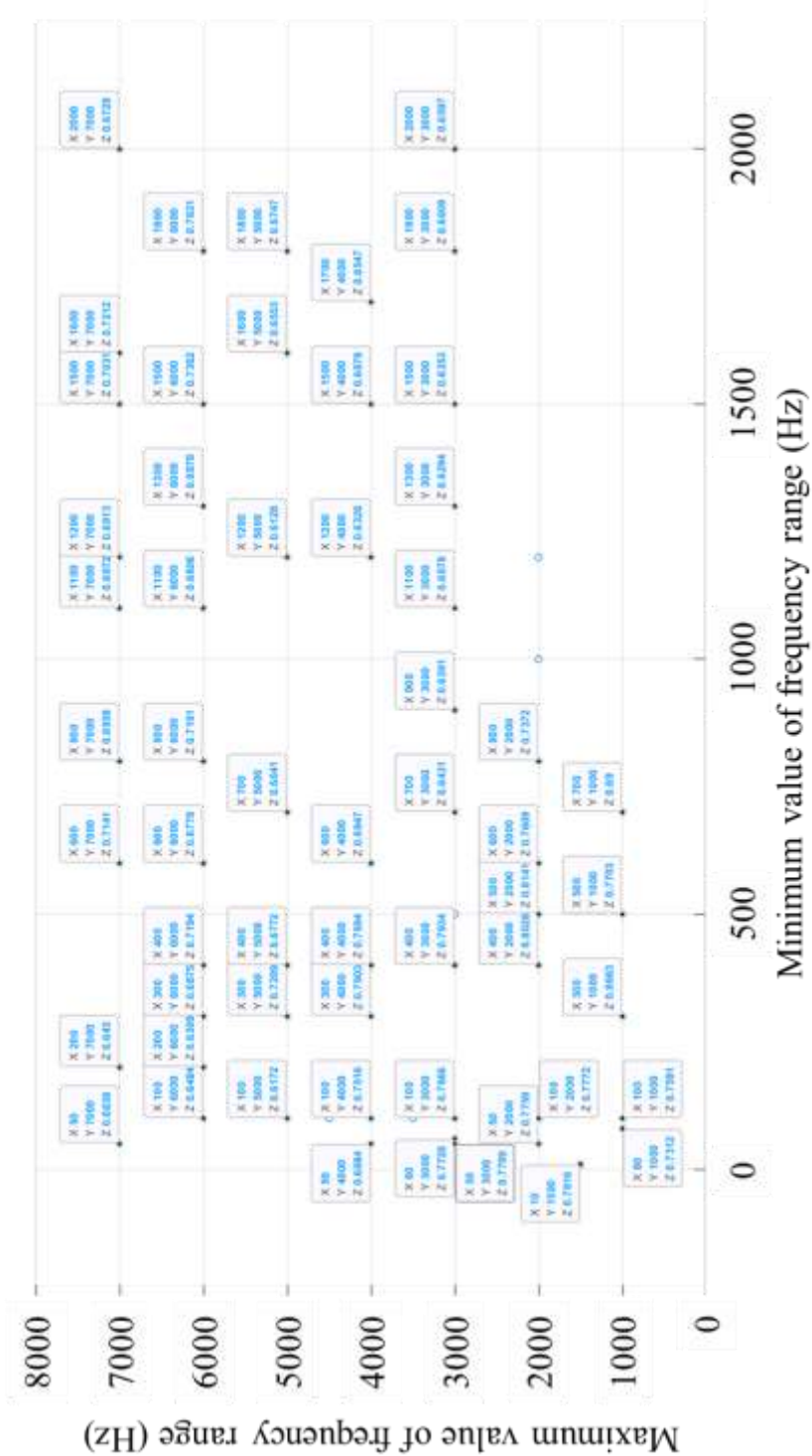


Figure 32 Monitoring accuracy with different frequency range

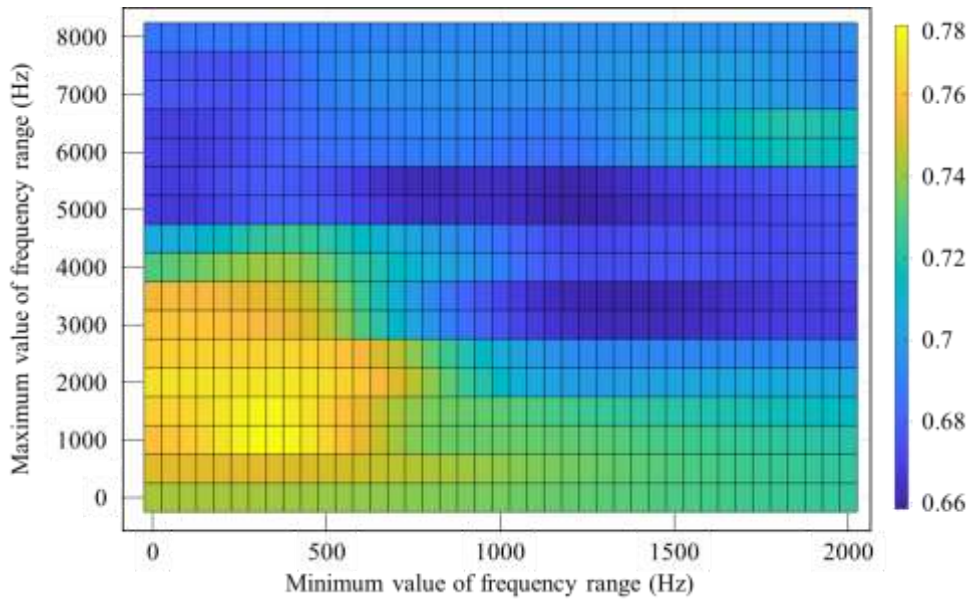
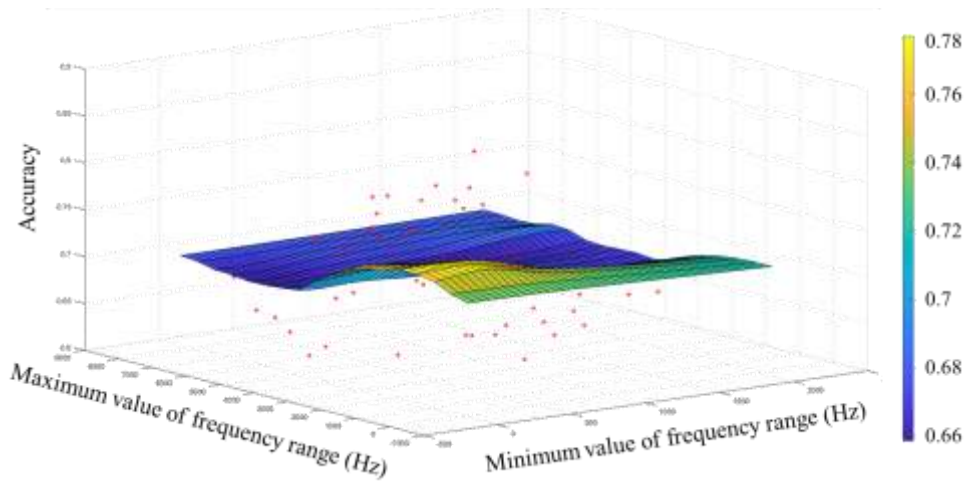


Figure 33 Estimated accuracy with various frequency range

### **3.4.2 Enhancement of performance: Various neuron network**

As mentioned in Section 2.7, this monitoring system uses a method of converting 1D data of 16,000 Hz based on time–magnitude into time–frequency–magnitude–based 2D data using Fourier Transform and Log–mel spectrogram, which is classified using artificial neural networks. Artificial neural networks play a role in classifying 2D data, and they use simple CNN like Figure 19, but it is safe to use other artificial neural networks that classify existing images. So in this section, Comparing performance was executed using other neural networks with more complex structures that are mainly used for image classification. The neural networks used here are simple CNN, ResNet–19, ResNet–50 and GoogLeNet.

Each neural network was used for bandaw monitoring among standard files and the target frequency domain was 50 Hz to 7000 Hz.

As Figure 34 shows, the monitoring performance of each network was 71%, 77%, 77%, and 79%, respectively, confirming that more complex networks could be improved, but not significantly.

Rather, considering the computational speed and learning volume of each network, and the size of the network, we could see the importance of selecting the appropriate network.

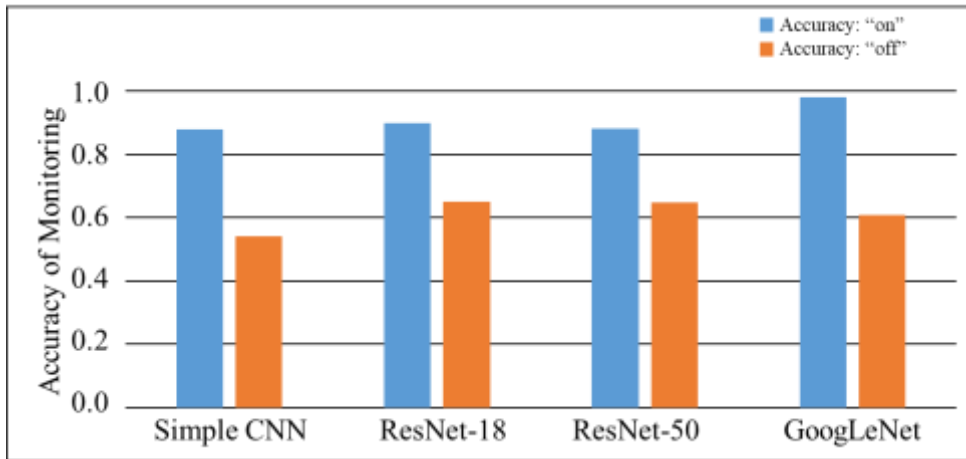


Figure 34 Accuracy of bandsaw monitoring

with various neuron network

Table 7 Test result with various neuron network

Network	Size of network	Input data	Classify time (640 images)
Simple-CNN	0.5 MB	40x98x1 Value	8.0 sec
ResNet-18	40 MB	40x98x3 RGB image	21.9 sec
ResNet-50	86 MB	40x98x3 RGB image	39.8 sec
GoogLeNet	22 MB	40x98x3 RGB image	37.5 sec

### **3.4.3 Wavelet transform based monitoring**

In case of log-mel spectrogram, 1-D data was transformed with 2-D data (time/frequency - intensity) based on STFT. However nowadays wavelet transform is often used. The strong point of wavelet transform is flexible time-frequency band. Unlike STFT, which uses trigonometric functions which make all the bandwidth same, wavelet transform has short time-long frequency band at the high-frequency range and long time-short frequency band at the low-frequency range, so wavelet transform can take advantage of the signal's characteristics better. Although it has the advantage of being able to better understand the characteristics of signals, it also has the disadvantage of increasing computation.

Using wavelet transform, 1-D data can be transform to 2-D data as shown in Figure 35 and training with existing simple CNN structure was executed. These trained network shows the 74% accuracy. It means just using wavelet transform performance can be improved from 70% to 75%. And these wavelet transformed data also trained with ResNet-50, more complex neural network, and recognition performance can be improved to 82.5% Detail result is as shown at Figure 36

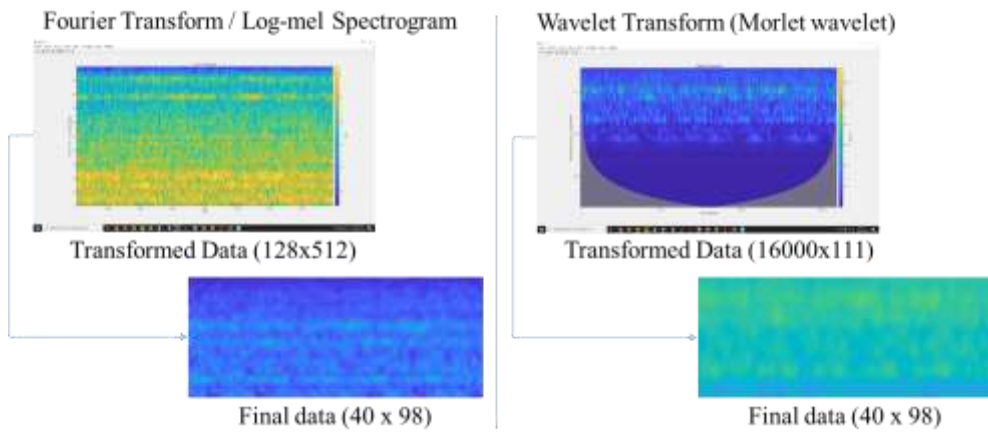


Figure 35 STFT based data process and wavelet transform based data process

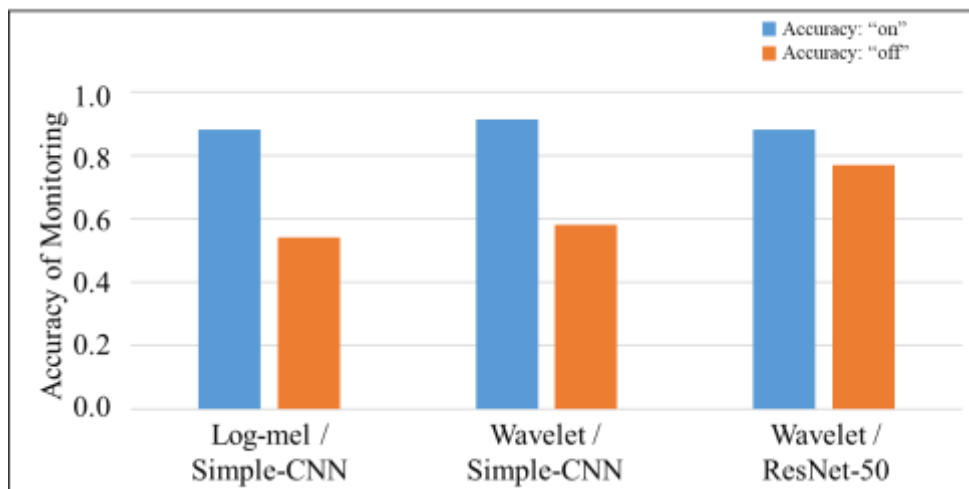


Figure 36 Recognition performance with wavelet transform

### **3.5 Virtual data set**

Sounds have the characteristic of being able to preserve their own information independently while overlapped. This means that it is possible to create a new virtual data set with combined, simple, mechanical sounds by combining .wav files. The advantages are that any size data set can be created, a data set can be created for a situation that is difficult to realize, and it is easy produce a new data set because it requires only recording the machine's operating sound rather than reproducing multiple situations.

Figure 37 shows the process of creating a virtual data set and the factors to consider. Using this process, virtual data sets were made for various situations by combining the operating sounds of the machines that we wanted to monitor.



(m: no. of devices to be monitored)

Process 1) Pooling the sample operation sound -

(p: no. of sample sound (Augmentation method 1))

$$P_i = \{D_1, D_2, \dots, D_p\} \quad (i = 1, 2, \dots, m) \dots\dots\dots(14)$$

(D<sub>i</sub>=Sample operation sound of device i, 16000Hz, 1sec)

Process 2) Extracting the sample from the pool

$$S_{pre-i} = \text{Rand}(P_i) \text{ or } \frac{\text{Rand}(P_i)}{\max(\text{Rand}(P_i))} \text{ or } \frac{\text{Rand}(P_i)}{\max(\text{Rand}(P_i))} \times r \dots\dots\dots(15)$$

(Rand(X): Randomly extract the elements from the set X)

(r<sub>i</sub>: Random value between 0 ~ 1)

Value of S<sub>i</sub> can be decided by setting of virtual data composition

Process 3) Cutting the extracted sample and paste

$$S_i(x) = S_{pre-i}(x + b - a) \quad (x = a, a + 1, \dots, a + s_i - 1)$$

$$S_i(x) = 0 \quad (x = 1, 2, \dots, a - 1, a + s_i, a + s_i + 1, \dots, 16000) \dots\dots\dots(16)$$

s<sub>i</sub>: random integer between 1 and 16000

(size of extraction)

a,b: random integer between 1 and (16000- s<sub>i</sub>)

(position of extraction and paste)

Process 4) Guide generation

( $g_{v,i}$ : Random value between “0 ~ 1” )

( $g_m$ : Random value between “0 ~  $(2^m - 1)$ ” or “1 ~  $(2^m - 1)$ ” )

Via binarizing the  $g_m$ , value of  $g_{m,i}$  can be defined. If you want to use the data with the zero sound, value of  $g_m$  can be define with random value between “0 ~  $(2^m - 1)$ ” , but if you do not want, , value of  $g_m$  can be define with random value between “1 ~  $(2^m - 1)$ ” .

Process 5) Composition of extracted data

( $j$ : no. of data at the virtual data set)

$$D_i = \{C_i | C_i = C_{p,i} \text{ or } \frac{C_{p,i}}{\max(C_{p,i})} \text{ or } \frac{C_{p,i}}{\max(C_{p,i})} \times r\} \quad (i = 1,2,3, \dots, j)$$

when  $C_{p,i} = \sum_{i=1}^m g_{v,i} \times g_{m,i} \times S_i \quad (i = 1,2, \dots, j) \dots\dots\dots(17)$

( $r$ : Random value between 0 ~ 1)

Value of  $C_i$  can be decided by setting of virtual data composition

Process 6) Adding extra factors like background noise

$$V_i = \{A_i | A_i = C_i + B_i\} \quad (i = 1,2,3, \dots, j) \dots\dots\dots(18)$$

( $B_i$ : Background noise data)

The following items were considered in the process:

(Factors 1)

The number of samples of data in the sound pool

(Factors 2)

Whether to modify the intensity of the extracted samples:

no modification, modification to the same intensity, modification to a random intensity

(Factor 3)

Length of sound to be cut and position to be paste

(Factors 4)

Whether to include a sample with no sound (all device off)

(Factors 5)

Whether to modify the intensity of the combined sample: no modification, modification to the same intensity, modification to a random intensity

(Factors 6)

Whether to include a background sound.

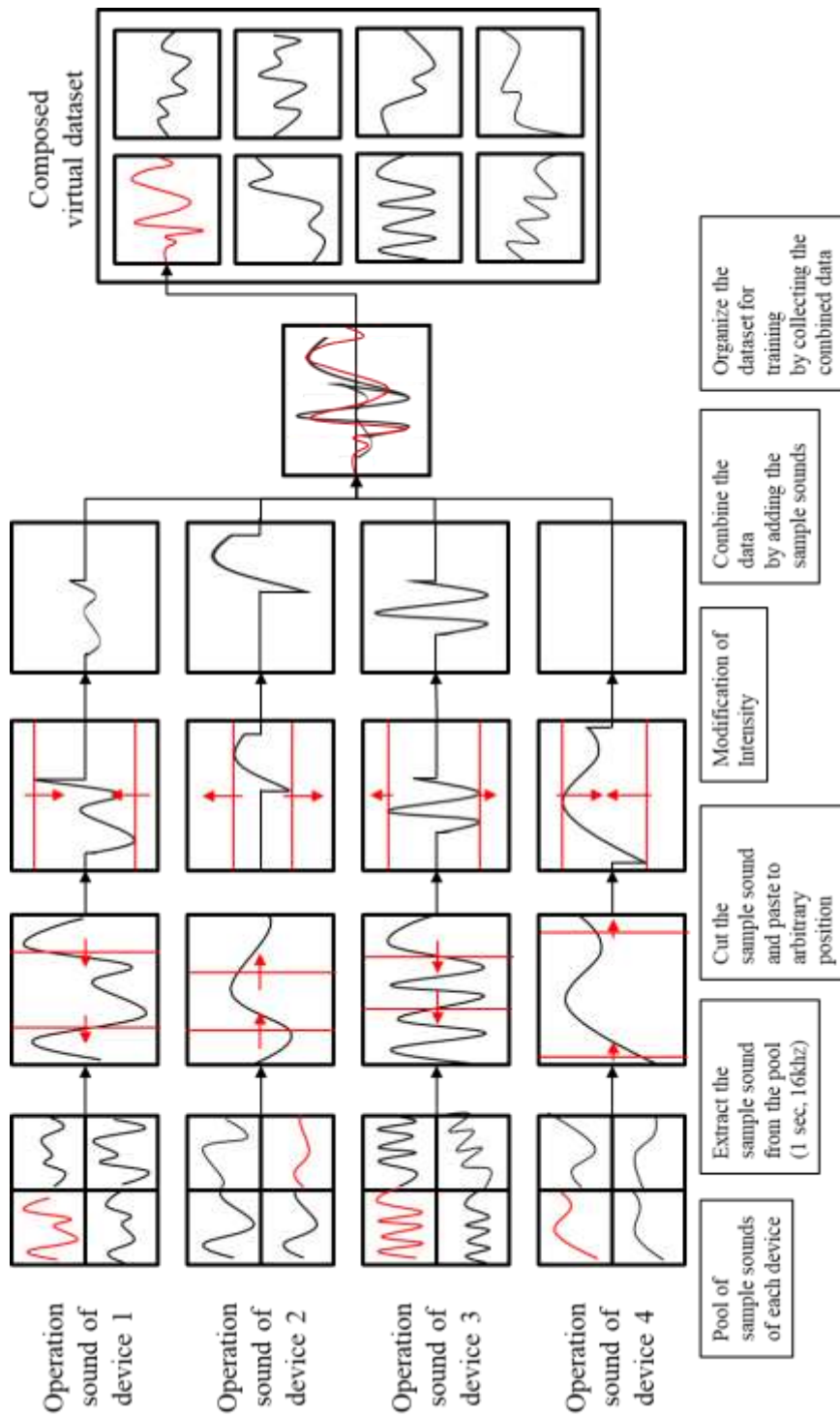


Figure 37 Schematic of the creation of a virtual data set for training

The performance of the monitoring system using the virtual data set under the conditions listed in Table 8 is as shown in Figure 38. The accuracy of the monitoring system was approximately 87%, 59%, 97%, and 99% for the bandsaw, drill, pump and turning, respectively. Thus, the monitoring system trained with the virtual data set operated just as it had when it was trained with real recordings. The performance for some devices improved by about 10%. However, for the drill, the monitoring performance was poor even with the virtual data set.

Table 8 Specification of the virtual dataset

Specification	Value
Raw data in the sample pool	10 files per device (1 sec per file)
Total data	2,999 .wav files
Intensity of raw data	Modified to the same intensity
Zero-sound sample	Excluded
Intensity of combined data	No modification
Background sound	Recorded sound when all devices are off
Frequency range	10 - 1,500 Hz (bandsaw) 50 - 7,000 Hz (other devices)

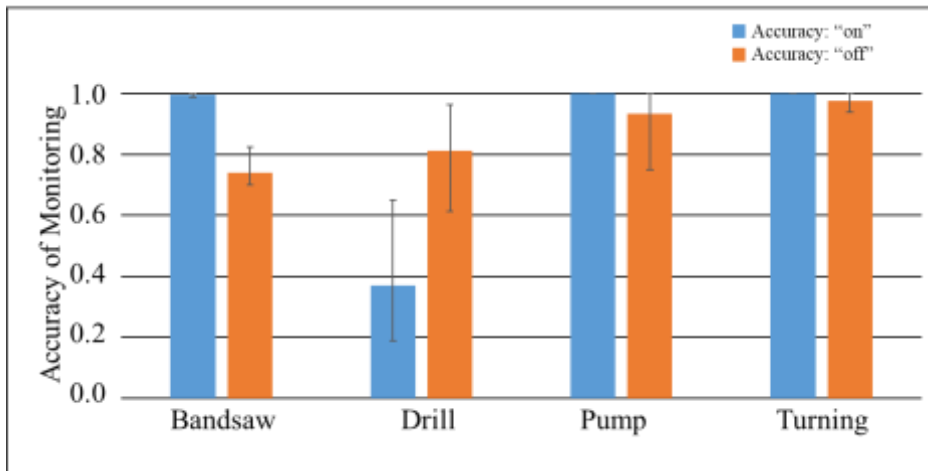


Figure 38 Performance evaluation of monitoring system based on virtual data set

### **3.6 Measuring intensity of sound based on masking**

In the case of operating sound from a mechanical device, signals are produced across the wide frequency bands, not in narrow frequency bands, which often have waveforms similar to noise, so it is not easy to separate sounds in the traditional way of separating sounds such as TDOA, beamforming, etc. Therefore, there is a limit to measuring the intensity of a particular sound.

In this section, measuring the intensity of operation sound from specific devices by applying masking to the algorithm that recognizes sound will be tried.

The principle used in this attempt is the overlay of sound. In other words, it is to manufacture a mask that reverses the sound of the machine that is intended to obtain intensity, and apply the mask from a weak point to a strong point to check whether the sound is recognized. When mask is applied and the monitoring system recognizes that there is no more operating sound of a particular equipment in the sound applied, the applied mask strength is defined as the intensity of the sound of specific devices. The detailed algorithms are as in Figure 39.

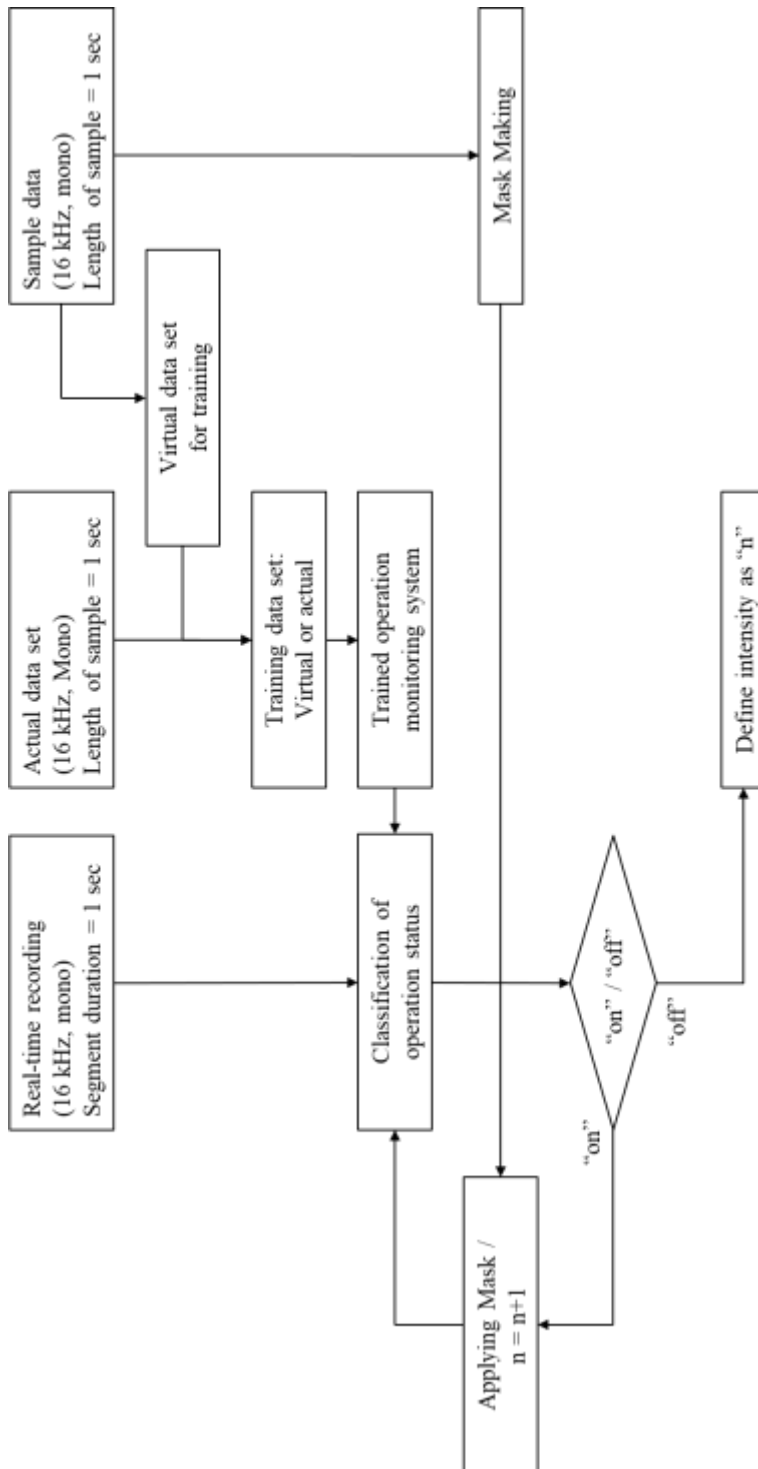


Figure 39 Concept diagram of intensity measuring



Method of producing Mask is as shown in Figure 40, and the detailed formula for this is as follows.

$$M(b) = \frac{(\max(M_p(b)) - M_p(b))}{N} \quad (M_p(b) = \sum_{t=0}^{t=1/f} (S(t \times f, b) \times f)) \quad \dots\dots\dots(19)$$

Relative intensity (I) can be expressed as

$$I = \frac{n}{N} \quad \dots\dots\dots(20)$$

M(b): mask

S(t,b): Intensity value from log-mel spectrogram

t: time

b: frequency band

f: frame length

N: mask division rate

n: no. of mask applying until trained net classified as “off”

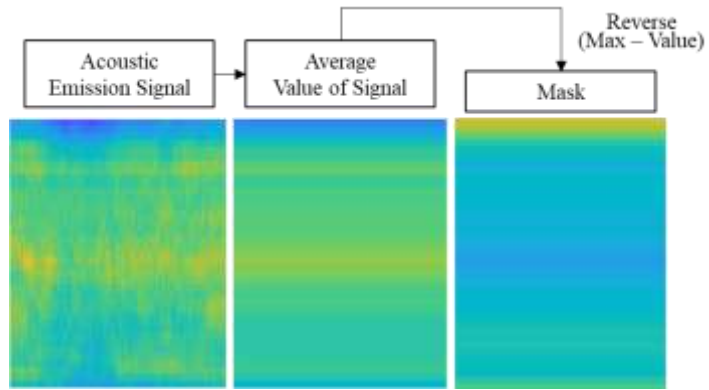


Figure 40 Fabrication of mask using existed operation sound data

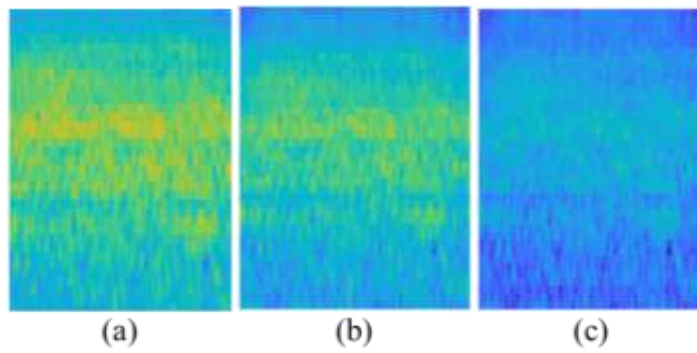


Figure 41 Applying mask with intensity value (a) 0 (b) 0.2 (c) 0.5

Measuring the intensity of pump operation sound using mask was tried. Since it is impossible to control the sound intensity of the pump itself, the experimenter took the mic array and moved away from the pump, then moved back to the pump, and measured the strength of the signal received by the mic array. Since the strength of sound is inversely proportional to the square of distance, the distance from the sound source pump and the strength of the sound are directly related. As can be seen in Figure 42, it was possible to confirm that the strength of the sound varies with the distance from the pump.

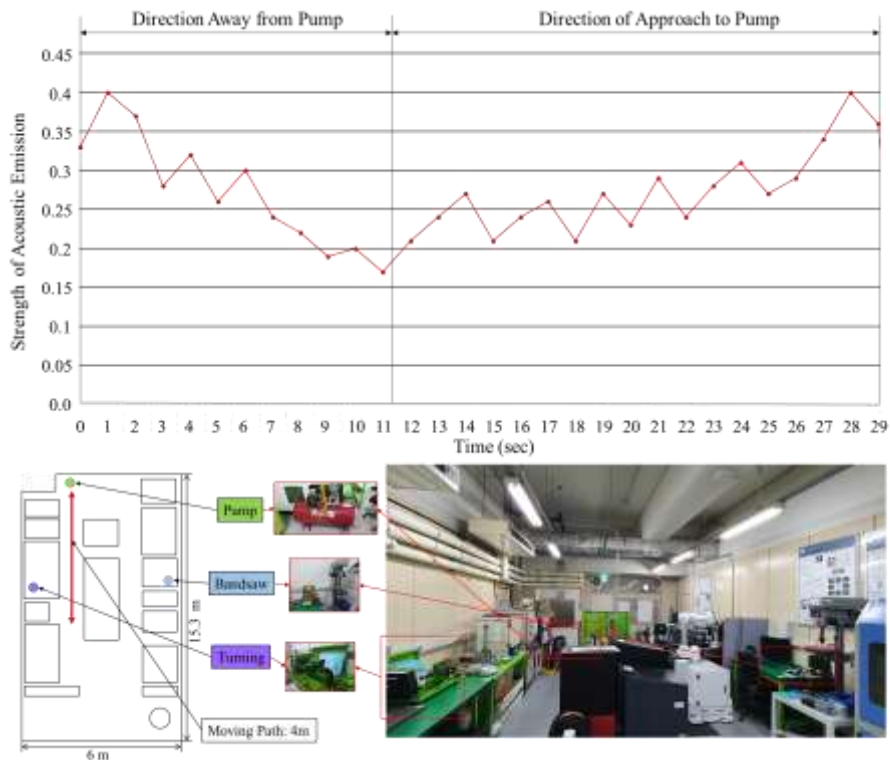


Figure 42 Measure the intensity of sound with different distance

Starting from the front of the pump, the change of sound was measured by moving one step at a time. At each step, mic array stayed for at least 5 seconds, and the distance between each step and step was specified at 0.45 m.

There was some error at the measured intensity at each step, but basically it was confirmed that when mic array move further from source, signal strength perceived by the mic array was getting smaller.

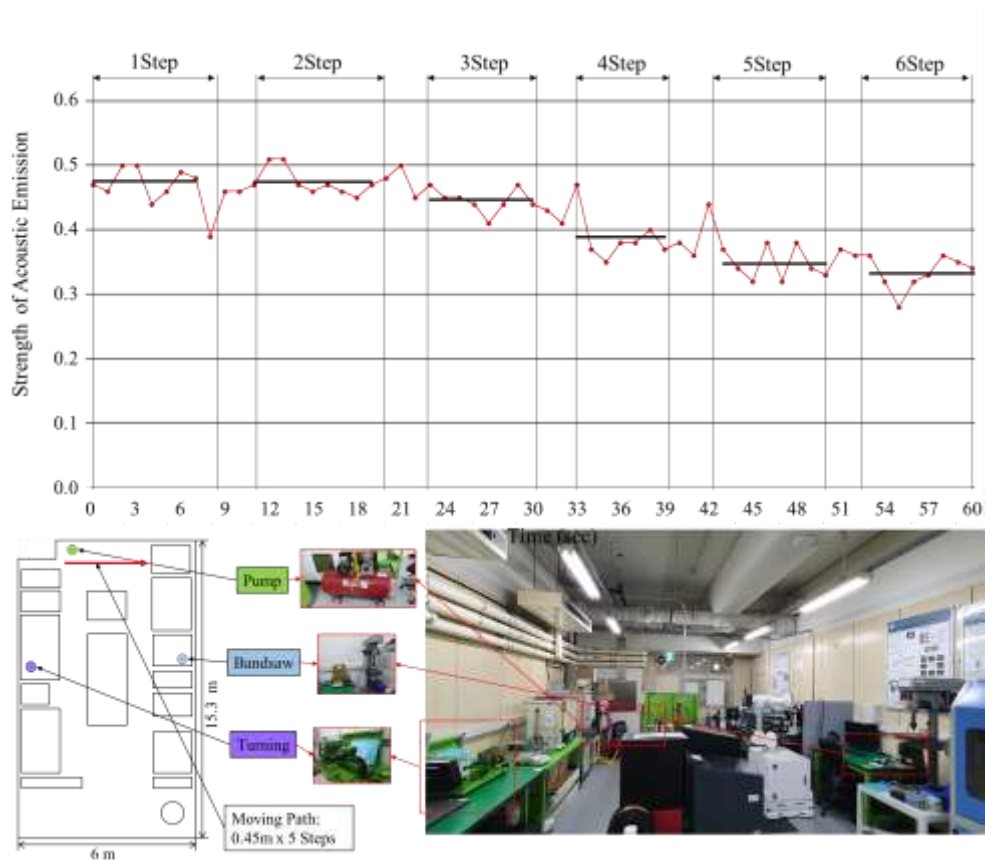


Figure 43 Measure the intensity of sound: step test

In order to verify the performance even operation sound of other devices was interrupted, the experiment was conducted while turning and bandsaw operating at the same time and the results were the same as Figure 44. As can be seen in Figure 44 regardless of whether turning and bandsaw were operated, recognized intensity of the pump sound was confirmed to be higher as the distance smaller and at the moment the pump operation was stopped, the perceived sound intensity was zero, even though other equipment was operated and emitted sound.

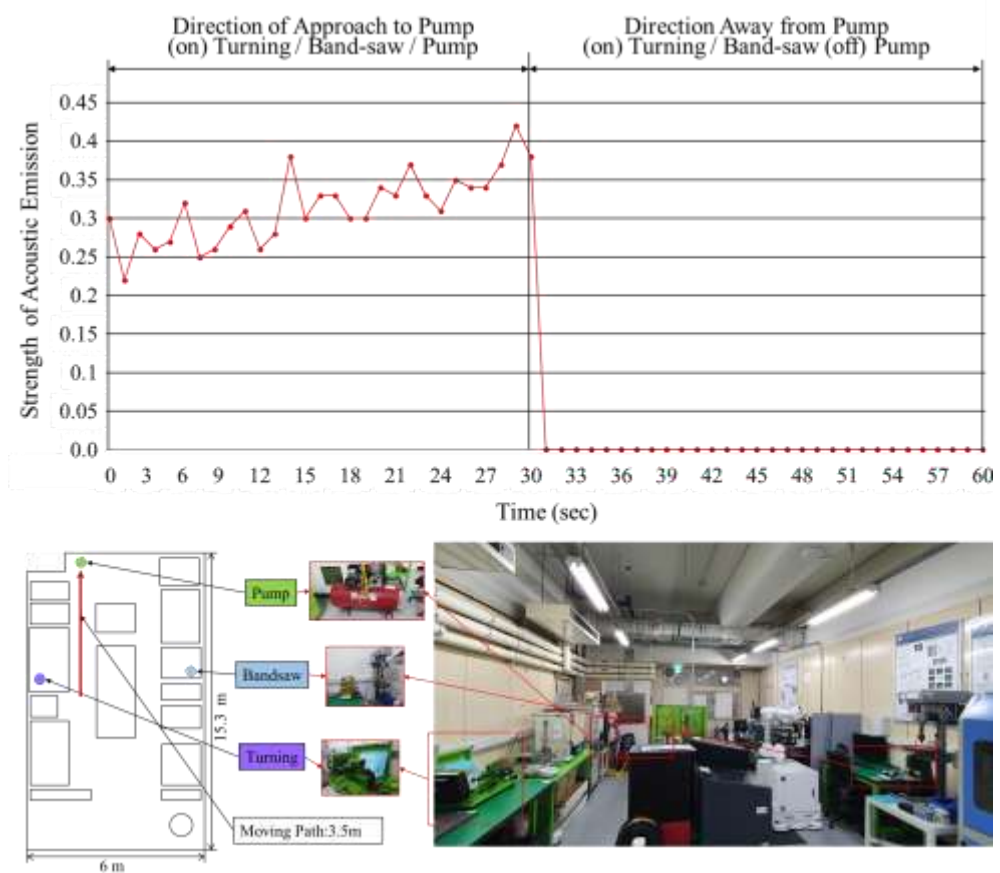


Figure 44 Measure the intensity of sound with different devices

## **Chap 4. Applying to real factory**

To verify the performance of the currently developed monitoring system, the application was attempted in two environments similar to the actual factory, not in the laboratory environment.

Because the system monitors the manufacturing process by determining whether a device is being used, it was decided that the system should be conservative and indicate only when device was definitely being used. As can be seen at the Figure 45, due to the characteristics of CNN, phase can be divided with “clearly predicted section” and “not clearly predicted section”, and most errors appear in “not clearly predicted section”. To guarantee the robustness, based on the predicted results obtained by the multi mic array, it was decided to exclude errors by indicating only when the average value of predicted probability from multi mic array that a device was on was “limit value” or higher

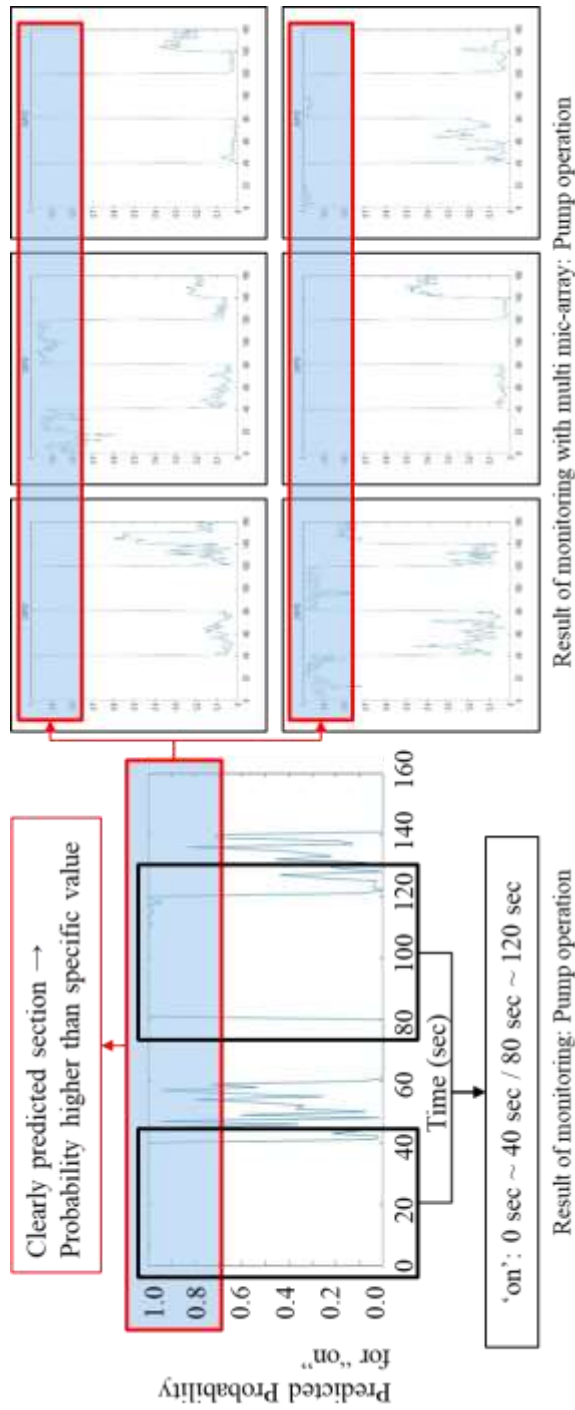


Figure 45 Clearly predicted section / not clearly predicted section

## **4.1 Case – Workshop with hand-operated device**

The first case of real monitoring was performed in a workshop with hand-operated device. The workshop was a space where students could freely create their own prototypes, and it was characterized by a variety of noises, including a fan for air conditioning.

The target devices were a turning, a bandsaw, and an airgun. All device had no electrical components except a simple power switch, and none of the device were smart devices. In this experiment, the air gun was the focus because of the very loud and unusual sound it emits and the fact that it is typically used between process steps for removing burrs or cleaning parts.





Figure 46 Workshop with hand-operated device

Table 9 Detail information about monitoring

Specification	Value
Monitoring period	800 sec
Neuron network	Simple CNN
Limit value	0.9
Dataset pool	90 sec per devices
Size of virtual dataset	9,999 .wav files

The monitoring results are shown in Figure 47. As can be seen, the operational status of each device was monitored accurately. 3 devices were monitored with high accuracy, 98% ~ 100%. In particular, it was shown that it was possible to infer the actual operation time of the machine in general, and that it was also possible to use this result and monitoring system to monitor the entire process.

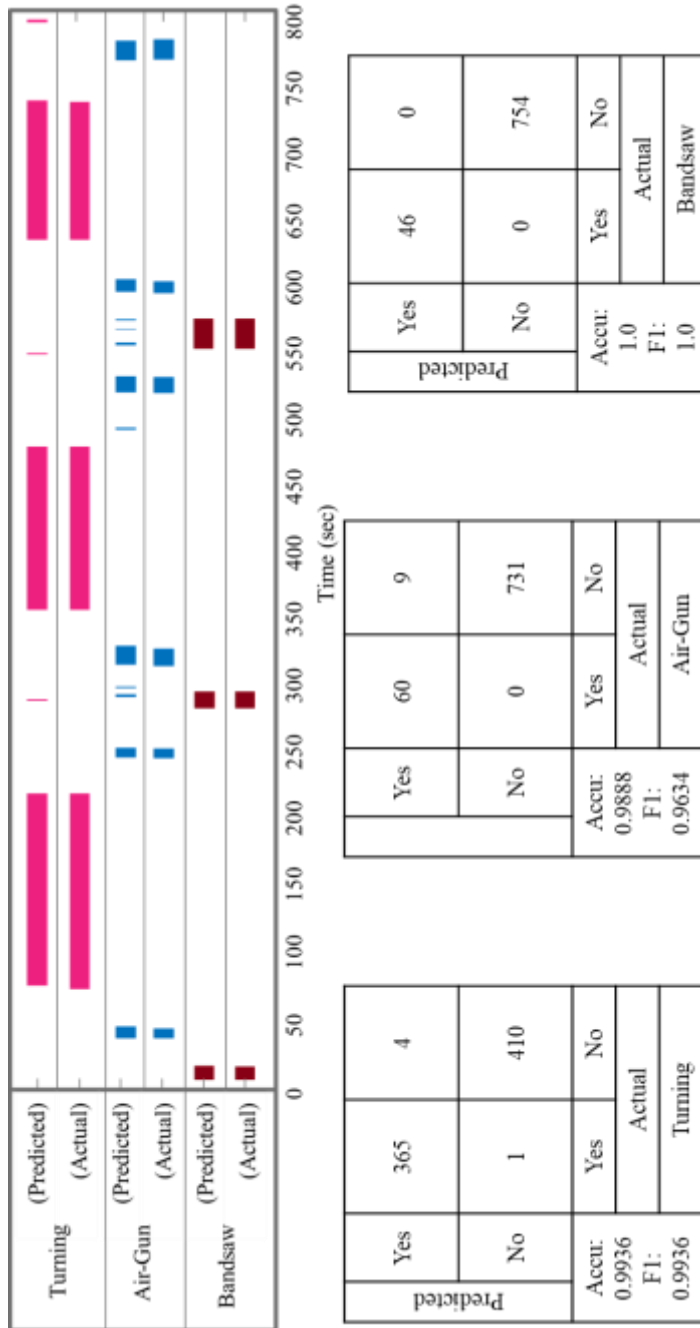


Figure 47 Monitoring results in a workshop with hand-operated devices

As a result of monitoring, it was possible to determine when the actual equipment was in use. Figure 48 is a result of analyzing how each process of this manufacturing process was conducted based on the monitoring results. As can be seen, it showed that even for hand-operated devices that do not have any IoT devices, the work can be linked to a central system and computerized. In other words, using the developed monitoring system, existing equipment could also be "connected" and "smart device" using mic array installed outside the equipment without extra equipment installation.

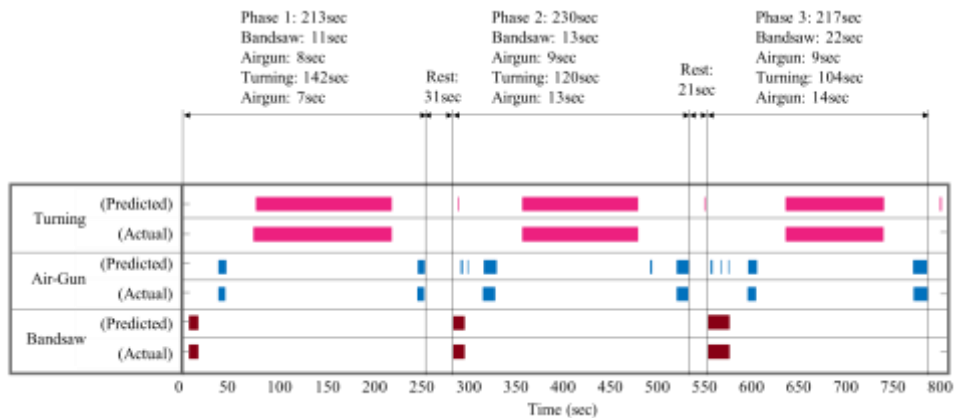


Figure 48 Process monitoring result of workshop with hand-operated devices

## **4.2 Case – Factory with a Computer Numerical Control (CNC) machine**

The other case of real monitoring was conducted at a factory that makes actual products using computer numerical control (CNC) machines, which are most commonly used in product manufacturing. The entire monitored CNC machine was controlled by a computer with firmware. It was impossible to measure and transmit the status of the device because no IoT-related functions could be performed separately outside of the machine's own system. In addition, because of the characteristics of the firmware provided by the manufacturer, it was impossible for the user to customize it to identify or transmit the status of the device.



Figure 49 Factory with a Computer Numerical Control (CNC) machine

Table 10 Detail information about monitoring

Specification	Value
Monitoring period	1,524 sec
Neuron network	ResNet-18
Limit value	0.7
Dataset pool	90 sec per devices
Size of virtual dataset	2,999 .wav files

The results of monitoring are shown in Figure 50. As can be seen, the operational status of each device was monitored accurately. 3 devices were monitored with high accuracy, 98%. The process being performed was drilling, but it was impossible to monitor the use of this device because it emitted no sound. However, it was possible to monitor the CNC machine's Automatic Tool Changer (ATC), which only operated when the machine was operating. There were mis-prediction between ATC and airgun due to the similarity of operation sound of these devices.

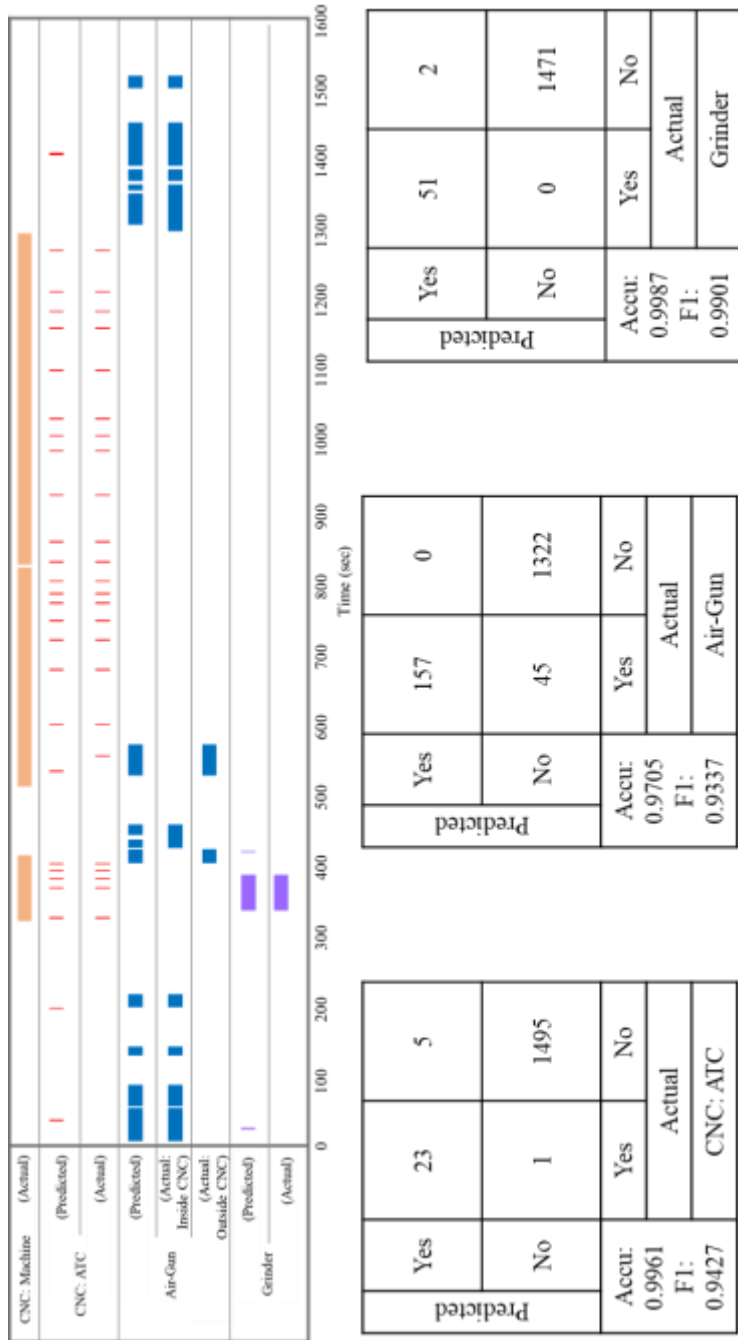


Figure 50 Monitoring results at a factory with a computer numerical control (CNC) machine



Due to the nature of the process, CNC Machine itself is a process in which the sound is very small compared to the ambient noise, so it is impossible to monitor with sound alone. ATC and airgun were used to monitor this indirectly. ATC is a device that moves when the CNC is operated. However, due to the characteristics of ATC, it is an instrument that operates eventfully during the process, so there is a limit to knowing the entire CNC machine operation time. In addition, there is no guarantee that the CNC will operate between the time and time when the ATC is activated.

So the device that I focused most on during this process monitoring is airgun. airgun is a device used mainly for cleaning. In order to carry out the process accurately, it must be operated before and after CNC operation. This means that the CNC will not run while the airgun is running.

Both ATC and airgun are devices that utilize compressed air, and have similar operating sounds. Thus, errors may occur in the recognition of mechanical sounds in both equipment. However, due to the characteristics of ATC that operates only during CNC operation and the characteristics of Airgun that operates only when CNC is not operated, if the two equipment is judged to have been operated at the same time, error handling was considered as an error.

(However, the airgun may be used externally for cleaning other devices while using CNC equipment, but this is to be treated as an exception.)

If it is determined that the airgun and ATC are operated simultaneously,

1) Short use within one or two seconds is considered ATC, considering the characteristics of the airgun that is not used for a short period like one or two seconds.

2) If the use is carried out in a long section for more than 3 seconds, it is considered as airgun has been operated, and if it is determined to be ATC operation during that time, it shall be deemed that it has not been operated.

This is used to infer the hours of use of CNC equipment as shown below.

1) If the airgun operated between the ATC operation hours, the CNC equipment is assumed to have not been operated.

2) If the airgun was not operated between the ATC operation hours, the CNC equipment is estimated to have been operated.

3) However, if there was a gap of more than 100 seconds between the ATC operating hours, the draft assumes that the CNC equipment was not activated.

The CNC operating time estimates in the light of the above principles are as shown in Figure 51. The CNC operation itself was not monitored, but CNC operating time can be predicted with an accuracy of about 92%.

It also succeeded in identifying how much time each process of the corresponding manufacturing process is consuming.

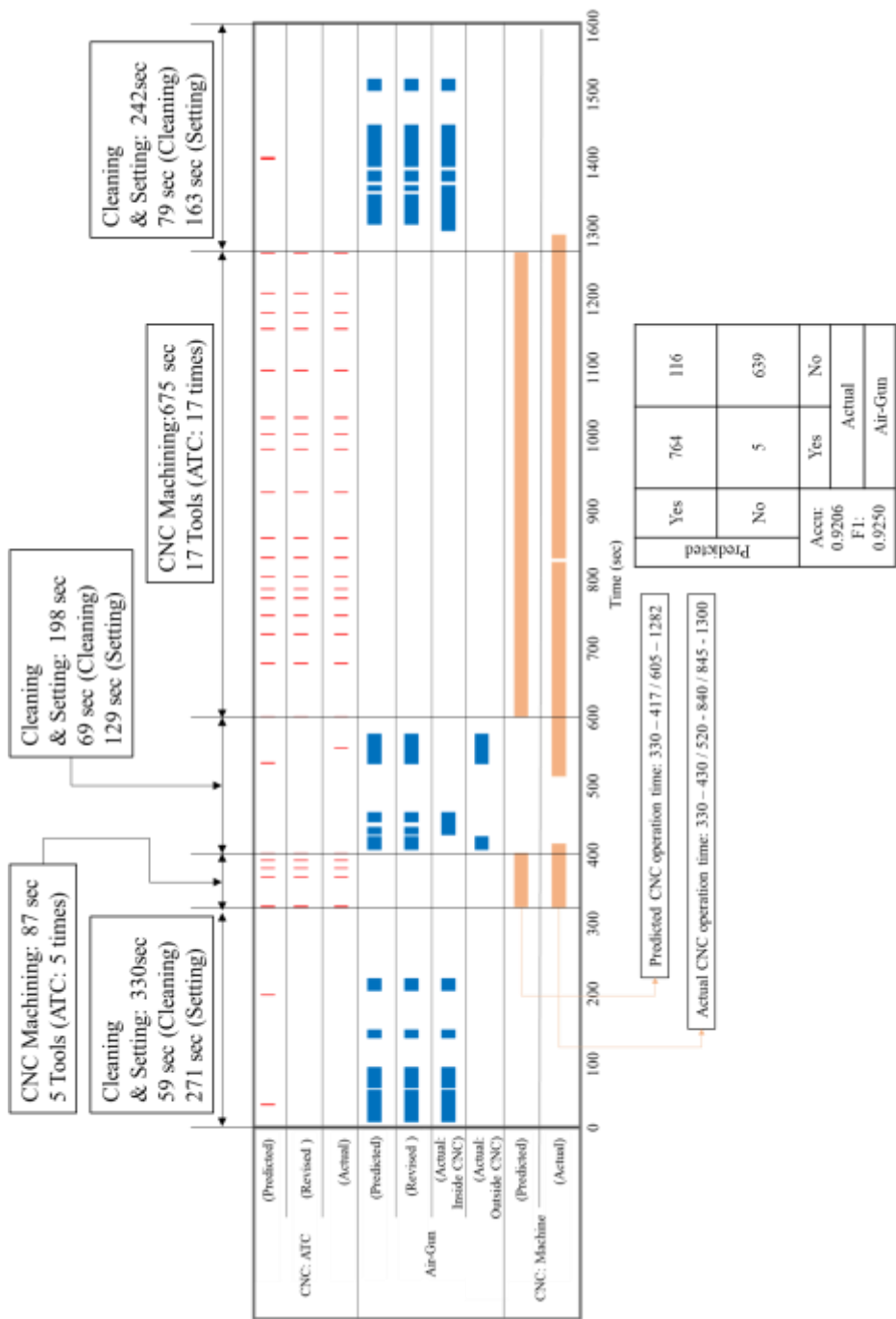


Figure 51 Process monitoring result of a factory with a computer numerical control (CNC) machine

### **4.3 Case – Factory with aluminum casting process**

The third monitoring site was the aluminum casting process line. The monitoring target is three equipment belonging to the line, the first is a cooler that cools down the heated die for casting, the second is a grinder that removes the burr of manufactured products, and the last is a separator that separates the handle of the product from the actual product.

Not only did the plant operate several very loud equipment at the same time, but it was also a site where a lot of noise was generated due to air conditioning equipment and various working noises.

The equipment of the plant was very old, and there were no IoT functions except simple control panel inside the machine. In addition, the factory's methods of work were all dependent on people, so simple tasks such as checking the number of works were all carried out manually, and simple analyses such as checking the hours of operation of equipment could not be performed.



Figure 52 Factory with aluminum casting process

Table 11 Detail information about monitoring

Specification	Value
Monitoring period	2,000 sec
Neuron network	Simple CNN
Limit value	0.8
Dataset pool	90 sec per devices
Size of virtual dataset	9,999 .wav files

The operation time of each equipment installed on the manufacturing line was as shown in Figure 53. As can be seen in the Figure 53, it can be confirmed that each equipment is operated repeatably with a constant cycle. It was monitored that after only the cooler operated six times alone, all the equipment continued to operate repeatedly with a constant cycle. Approximate operating hours were monitored using this monitoring system, but not all operating hours were accurately captured. In case of separator with 1 second work, some errors occurred when some of the working time spanned by two frames separated by 1 second. From this, it was shown that consideration of the frequency and overall length of the data was needed. However, it was possible to determine roughly whether and how often the equipment was operating.

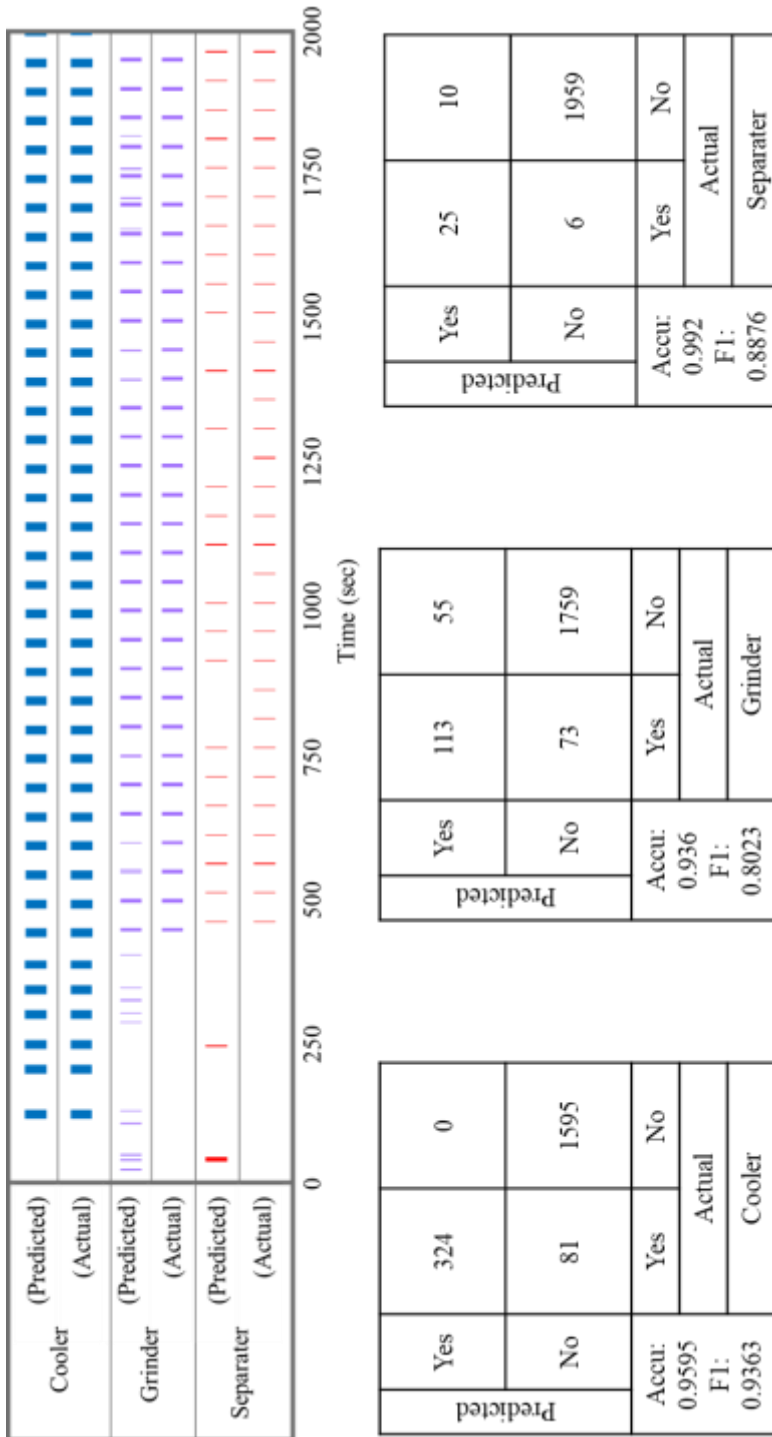


Figure 53 Monitoring results

at a factory with aluminum casting process



Figure 54 is the result of monitoring the actual process based on equipment operation status. Monitoring was carried out from the break time (lunch time). Preliminary operation that was only with cooler operation and test fabrication was executed. 6 times of preliminary operation was monitored as operation of cooler without the operation of grinder and separator. After preliminary operation, at the 13:21:09, main manufacturing process with operation of cooler / grinder / separator was started and this main manufacturing process was also successfully monitored. Exact number of processes could be counted and process time for each process can also be checked. This process data can be used as the data for process optimization in the future.

In this very noisy environment, for the actual product production process, this monitoring system has demonstrated that it can be successfully monitored.

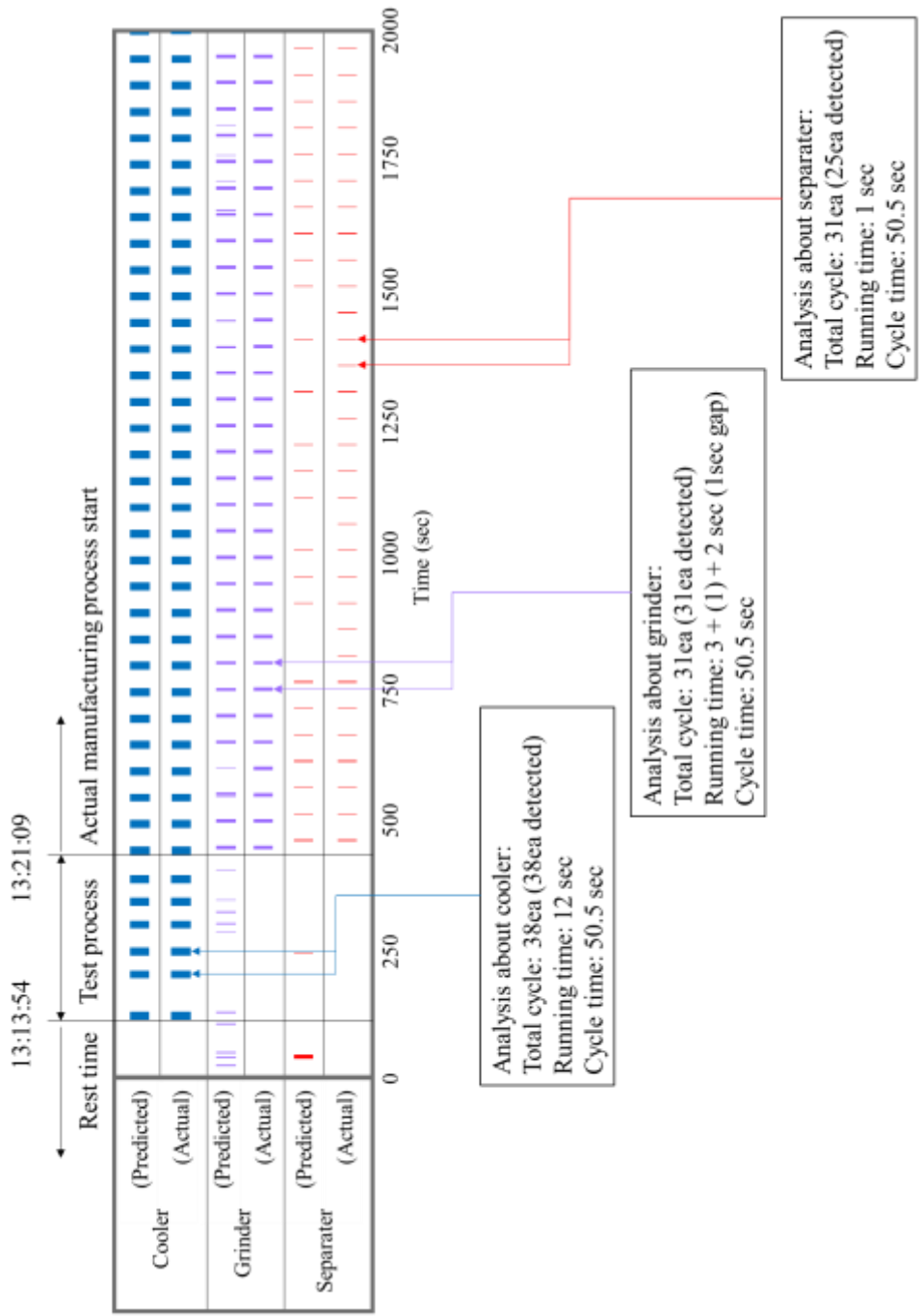


Figure 54 Process monitoring result of a factory with a factory with aluminum casting process

## **Chap 5 Application**

### **5.1 Application: Sound-based manufacturing process monitoring system**

Application of this research is a sound-based comprehensive process monitoring technology. Install a mic array near the line where the process is being carried out, listen to the operation sound in real time, and upload it to the central cloud server using the Internet. It is also possible to analyze the uploaded voice signal to identify which equipment is currently operating, to store the history of the operation, to determine how the equipment was operated, and to monitor the entire process. Especially for this purpose, it is sufficient to install only microarray near the equipment without having to install a separate device on the equipment itself, so it is possible to make smart devices for outdated equipment at low cost, obtain process information, and conduct separate research.

Figure 55 is the example of sound based manufacturing process monitoring system.

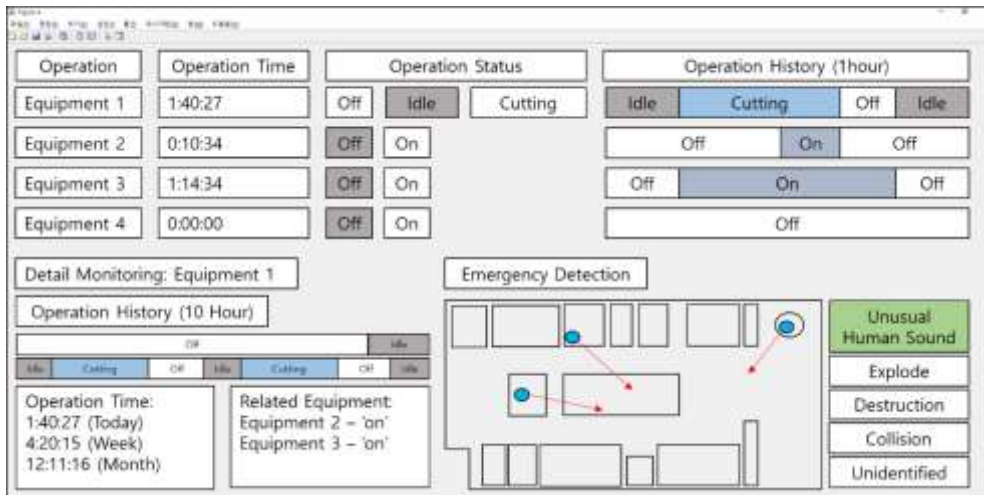


Figure 55 Example of Sound-based manufacturing process monitoring system

Monitoring the operation of a CNC machine's ATC showed that it is possible to monitor events based on instantaneous sound as well as continuous operating sound. Thus, further research is planned to improve the stability of the processes in a factory by continuously monitoring for various emergency situations.

When using mic array, it is possible to derive the direction where sound occurred from by using the phase difference caused by Time Delay of Arrival (TDOA) caused by the space between each mic. Figure 56 shows the location of a three-dimensional mobile device (drone) by identifying the location of the sound source with the mic array used at the monitoring system. This means that the location of the sound source can be sufficiently determined by the using mic array, which shows that it is possible to recognize a particular sound that can occur in an emergency situation setup by combining multiple mic arrays and to track where it originated when the sound is recognized.

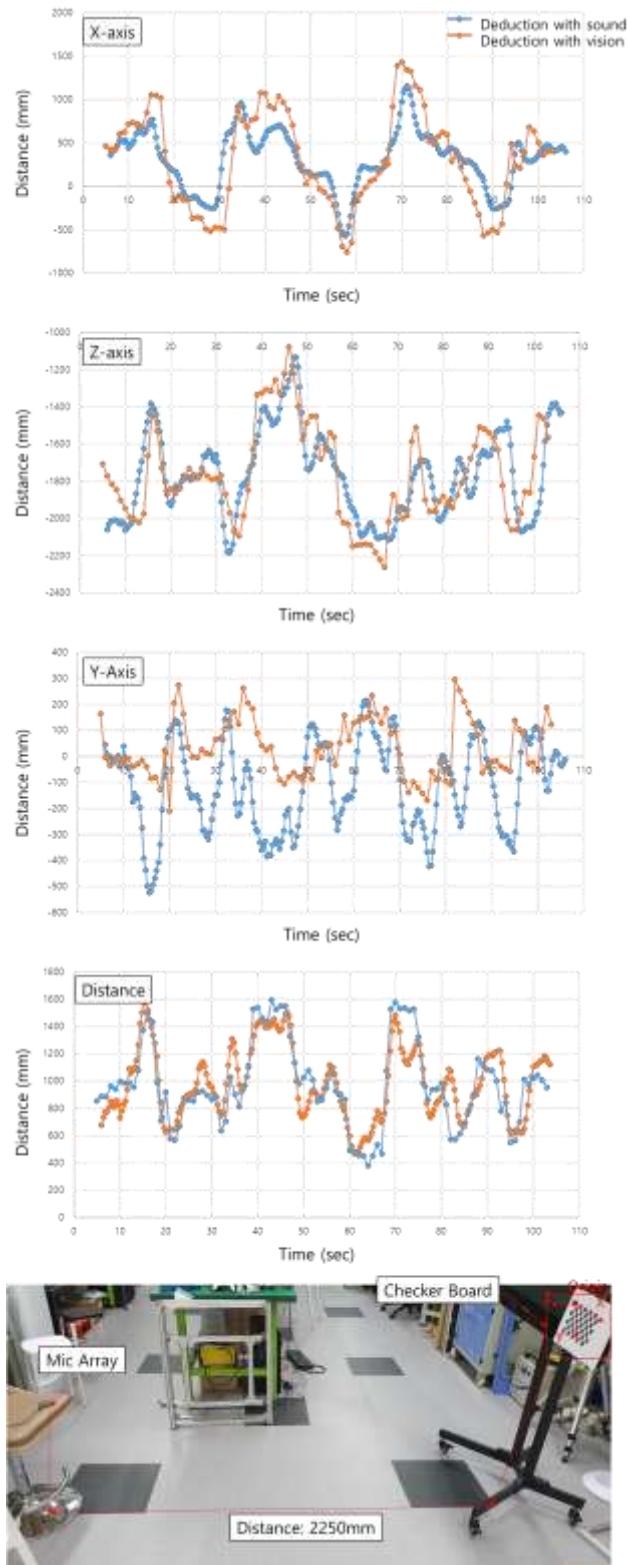


Figure 56 Drone position detection by sound analyzing

## **Chap 6 Conclusion**

Here a system for monitoring the operational status of multiple devices operating simultaneously by classifying their operational sounds using a CNN was developed.

When the CNN was trained with recorded operating sounds, the system recognized the operational status of the device with an accuracy of approximately 71-92%. However, it could not appropriately monitor quiet devices, such as a drill. After we modified the targeted frequency range, accuracy improved to 71% to 85% for the bandsaw. When trained with a virtual data set created by combining 1 sec sound files from each device, the system had an accuracy of approximately 87-99%.

To verify that the monitoring system worked without problems in an actual working environment, monitoring system was tested in a workshop and a factory. When the system was trained using only recorded sounds rather than a virtual data set, it detected the operational statuses of a devices only with the mic array at the environment with loud noise.

## 참고 문헌

- [1] K. Schwab, The Fourth Industrial Revolution: what it means, how to respond, World Economic Forum, 2016.
- [2] T. Kalsoom, N. Ramzan, S. Ahmed, M. Ur-Rehman, Advances in Sensor Technologies in the Era of Smart Factory and Industry 4.0, *Sensors* 20(23) (2020).
- [3] H.D. Morris, S. Ellis, J. Feblowitz, K. Knickle, M. Torchia, A Software Platform for Operational Technology Innovation, IDC White Paper, International Data Corporation (IDC), 2014.
- [4] J. Leng, Q. Chen, N. Mao, P. Jiang, Combining granular computing technique with deep learning for service planning under social manufacturing contexts, *Knowledge-Based Systems* 143 (2018) 295–306.
- [5] Y.G. Kim, S. Lee, J. Son, H. Bae, B.D. Chung, Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system, *Journal of Manufacturing Systems* 57 (2020) 440–450.
- [6] X. Shao, C.S. Kim, D.G. Kim, Accurate Multi-Scale Feature Fusion CNN for Time Series Classification in Smart Factory, *Computers, Materials & Continua* 65(1) (2020).
- [7] J. Wang, Y. Ma, L. Zhang, R.X. Gao, D. Wu, Deep learning for smart manufacturing: Methods and applications, *Journal of Manufacturing Systems* 48 (2018) 144–156.
- [8] D. Peddireddy, X. Fu, H. Wang, B.G. Joung, V. Aggarwal, J.W. Sutherland, M. Byung-Guk Jun, Deep Learning Based Approach for Identifying Conventional Machining Processes from CAD Data, *Procedia Manufacturing* 48 (2020) 915–925.
- [9] J. Kim, H. Lee, S. Jeong, S.-H. Ahn, Sound-based remote real-time multi-device operational monitoring system using a Convolutional Neural Network (CNN), *Journal of Manufacturing Systems* 58 (2021) 431–441.
- [10] <https://www.salesforce.com/blog/2018/12/what-is-the-fourth-industrial-revolution-4IR.html>.
- [11] M.E. Peralta, V.M. Soltero, Analysis of fractal manufacturing systems framework towards industry 4.0, *Journal of Manufacturing Systems* 57 (2020) 46–60.
- [12] W.-S. Chu, M.-S. Kim, K.-H. Jang, J.-H. Song, H. Rodrigue, D.-M. Chun, Y.T. Cho, S.H. Ko, K.-J. Cho, S.W. Cha, S. Min, S.H. Jeong, H. Jeong, C.-M. Lee, C.N. Chu, S.-H. Ahn, From design for manufacturing (DFM) to manufacturing for design (MFD) via hybrid



manufacturing and smart factory: A review and perspective of paradigm shift, *International Journal of Precision Engineering and Manufacturing–Green Technology* 3(2) (2016) 209–222.

[13] K.–H. Büttner, U. Brück, Use Case Industrie 4.0–Fertigung im Siemens Elektronikwerk Amberg, *Handbuch Industrie 4.0 Bd.4* (2017) 45–70.

[14] J. Leng, G. Ruan, P. Jiang, K. Xu, Q. Liu, X. Zhou, C. Liu, Blockchain–empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey, *Renewable and Sustainable Energy Reviews* 132 (2020) 110112.

[15] J. Leng, P. Jiang, K. Xu, Q. Liu, J.L. Zhao, Y. Bian, R. Shi, Makerchain: A blockchain with chemical signature for self–organizing process in social manufacturing, *Journal of Cleaner Production* 234 (2019) 767–778.

[16] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, B. Yin, Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges, *IEEE Access* 6 (2018) 6505–6519.

[17] H. Sun, G. Pedrielli, G. Zhao, C. Zhou, W. Xu, R. Pan, Cyber coordinated simulation for distributed multi–stage additive manufacturing systems, *Journal of Manufacturing Systems* 57 (2020) 61–71.

[18] D.–H. Kim, T.J.Y. Kim, X. Wang, M. Kim, Y.–J. Quan, J.W. Oh, S.–H. Min, H. Kim, B. Bhandari, I. Yang, S.–H. Ahn, Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry, *International Journal of Precision Engineering and Manufacturing–Green Technology* 5(4) (2018) 555–568.

[19] J. Leng, D. Yan, Q. Liu, H. Zhang, G. Zhao, L. Wei, D. Zhang, A. Yu, X. Chen, Digital twin–driven joint optimisation of packing and storage assignment in large–scale automated high–rise warehouse product–service system, *International Journal of Computer Integrated Manufacturing* (2019) 1–18.

[20] J. Leng, H. Zhang, D. Yan, Q. Liu, X. Chen, D. Zhang, Digital twin–driven manufacturing cyber–physical system for parallel controlling of smart workshop, *Journal of Ambient Intelligence and Humanized Computing* 10(3) (2019) 1155–1166.

[21] P. Wang, M. Luo, A digital twin–based big data virtual and real fusion learning reference framework supported by industrial internet towards smart manufacturing, *Journal of Manufacturing Systems* 58 (2021) 16–32.

[22] J. Huang, Q. Chang, J. Arinez, Product Completion Time Prediction Using A Hybrid Approach Combining Deep Learning and System Model, *Journal of Manufacturing Systems* 57 (2020) 311–322.

- [23] D.G.S. Pivoto, L.F.F. de Almeida, R. da Rosa Righi, J.J.P.C. Rodrigues, A.B. Lugli, A.M. Alberti, Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review, *Journal of Manufacturing Systems* 58 (2021) 176–192.
- [24] K.Y.H. Lim, P. Zheng, C.-H. Chen, L. Huang, A digital twin-enhanced system for engineering product family design and optimization, *Journal of Manufacturing Systems* 57 (2020) 82–93.
- [25] T.J. Rato, M.S. Reis, An integrated multiresolution framework for quality prediction and process monitoring in batch processes, *Journal of Manufacturing Systems* 57 (2020) 198–216.
- [26] P. Fang, J. Yang, L. Zheng, R.Y. Zhong, Y. Jiang, Data analytics-enable production visibility for Cyber-Physical Production Systems, *Journal of Manufacturing Systems* 57 (2020) 242–253.
- [27] J. Lenz, E. MacDonald, R. Harik, T. Wuest, Optimizing smart manufacturing systems by extending the smart products paradigm to the beginning of life, *Journal of Manufacturing Systems* 57 (2020) 274–286.
- [28] J. Lee, Y.C. Lee, J.T. Kim, Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database, *Journal of Manufacturing Systems* 57 (2020) 357–366.
- [29] M. Debevec, M. Simic, V. Jovanovic, N. Herakovic, Virtual factory as a useful tool for improving production processes, *Journal of Manufacturing Systems* 57 (2020) 379–389.
- [30] N.-H. Tran, Park, H.-S, Nguyen, Q.-V, Hoang, T.-D., Development of a Smart Cyber-Physical Manufacturing System in the Industry 4.0 Context, *Appl. Sci* 9(16) (2019) 3325.
- [31] F. Shrouf, J. Ordieres, G. Miragliotta, Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm, 2014 IEEE International Conference on Industrial Engineering and Engineering Management, 2014, pp. 697–701.
- [32] J. Leng, D. Yan, Q. Liu, K. Xu, J.L. Zhao, R. Shi, L. Wei, D. Zhang, X. Chen, ManuChain: Combining Permissioned Blockchain With a Holistic Optimization Model as Bi-Level Intelligence for Smart Manufacturing, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50(1) (2020) 182–192.
- [33] K.-C. Ying, P. Pourhejazy, C.-Y. Cheng, C.-H. Wang, Cyber-physical assembly system-based optimization for robotic assembly sequence planning, *Journal of Manufacturing Systems* 58 (2021) 452–466.
- [34] C. Liebrecht, M. Kandler, M. Lang, S. Schaumann, N. Stricker, T. Wuest, G. Lanza, Decision support for the implementation of

Industry 4.0 methods: Toolbox, Assessment and Implementation Sequences for Industry 4.0, *Journal of Manufacturing Systems* 58 (2021) 412–430.

[35] E.S. Okpoti, I.–J. Jeong, A reactive decentralized coordination algorithm for event–driven production planning and control: A cyber–physical production system prototype case study, *Journal of Manufacturing Systems* 58 (2021) 143–158.

[36] J. Wang, Y. Li, R. Zhao, R.X. Gao, Physics guided neural network for machining tool wear prediction, *Journal of Manufacturing Systems* 57 (2020) 298–310.

[37] K. Zhang, J. Chen, T. Zhang, S. He, T. Pan, Z. Zhou, Intelligent fault diagnosis of mechanical equipment under varying working condition via iterative matching network augmented with selective Signal reuse strategy, *Journal of Manufacturing Systems* 57 (2020) 400–415.

[38] Q. Wang, W. Jiao, Y. Zhang, Deep learning–empowered digital twin for visualized weld joint growth monitoring and penetration control, *Journal of Manufacturing Systems* 57 (2020) 429–439.

[39] P. Lin, M. Li, X. Kong, J. Chen, G.Q. Huang, M. Wang, Synchronisation for smart factory – towards IoT–enabled mechanisms, *International Journal of Computer Integrated Manufacturing* 31(7) (2018) 624–635.

[40] J. Wan, J. Yang, Z. Wang, Q. Hua, Artificial Intelligence for Cloud–Assisted Smart Factory, *IEEE Access* 6 (2018) 55419–55430.

[41] K.T. Park, Y.T. Kang, S.G. Yang, W.B. Zhao, Y.–S. Kang, S.J. Im, D.H. Kim, S.Y. Choi, S. Do Noh, Cyber Physical Energy System for Saving Energy of the Dyeing Process with Industrial Internet of Things and Manufacturing Big Data, *International Journal of Precision Engineering and Manufacturing–Green Technology* 7(1) (2020) 219–238.

[42] H. Kim, W.–K. Jung, I.–G. Choi, S.–H. Ahn, A Low–Cost Vision–Based Monitoring of Computer Numerical Control (CNC) Machine Tools for Small and Medium–Sized Enterprises (SMEs), *Sensors* 19(20) (2019) 4506.

[43] J.M. Müller, K.–I. Voigt, Sustainable Industrial Value Creation in SMEs: A Comparison between Industry 4.0 and Made in China 2025, *International Journal of Precision Engineering and Manufacturing–Green Technology* 5(5) (2018) 659–670.

[44] W.–K. Jung, D.–R. Kim, H. Lee, T.–H. Lee, I. Yang, B.D. Youn, D. Zontar, M. Brockmann, C. Brecher, S.–H. Ahn, Appropriate Smart Factory for SMEs: Concept, Application and Perspective, *International Journal of Precision Engineering and Manufacturing*

(2020).

- [45] G.-Y. Lee, M. Kim, Y.-J. Quan, M.-S. Kim, T.J.Y. Kim, H.-S. Yoon, S. Min, D.-H. Kim, J.-W. Mun, J.W. Oh, I.G. Choi, C.-S. Kim, W.-S. Chu, J. Yang, B. Bhandari, C.-M. Lee, J.-B. Ihn, S.-H. Ahn, Machine health management in smart factory: A review, *Journal of Mechanical Science and Technology* 32(3) (2018) 987–1009.
- [46] E.S. Gademawla, Computer vision algorithms for measurement and inspection of external screw threads, *Measurement* 100 (2017) 36–49.
- [47] D.Y. Pimenov, A. Bustillo, T. Mikolajczyk, Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth, *Journal of Intelligent Manufacturing* 29(5) (2018) 1045–1061.
- [48] Y. Miao, J.Y. Jeon, G. Park, An image processing-based crack detection technique for pressed panel products, *Journal of Manufacturing Systems* 57 (2020) 287–297.
- [49] J.P. Yun, W.C. Shin, G. Koo, M.S. Kim, C. Lee, S.J. Lee, Automated defect inspection system for metal surfaces based on deep learning and data augmentation, *Journal of Manufacturing Systems* 55 (2020) 317–324.
- [50] T.P. Nguyen, S. Choi, S.-J. Park, S.H. Park, J. Yoon, Inspecting Method for Defective Casting Products with Convolutional Neural Network (CNN), *International Journal of Precision Engineering and Manufacturing-Green Technology* (2020).
- [51] W. Wu, Y. Zheng, K. Chen, X. Wang, N. Cao, A Visual Analytics Approach for Equipment Condition Monitoring in Smart Factories of Process Industry, 2018 IEEE Pacific Visualization Symposium (PacificVis), 2018, pp. 140–149.
- [52] M. Hayashi, H. Yoshioka, H. Shinno, An Adaptive Control of Ultraprecision Machining with an In-Process Micro-Sensor, *Journal of Advanced Mechanical Design, Systems, and Manufacturing* 2(3) (2008) 322–331.
- [53] A. Massaro, A. Galiano, G. Meuli, S. Massari, Overview and Application of Enabling Technologies Oriented on Energy Routing Monitoring, on Network Installation and on Predictive Maintenance, *International Journal of Artificial Intelligence & Applications* 9 (2018) 01–20.
- [54] P. Farahmand, R. Kovacevic, An experimental-numerical investigation of heat distribution and stress field in single- and multi-track laser cladding by a high-power direct diode laser, *Optics & Laser Technology* 63 (2014) 154–168.
- [55] S. Hu, F. Liu, Y. He, T. Hu, An on-line approach for energy efficiency monitoring of machine tools, *Journal of Cleaner Production*

27 (2012) 133–140.

[56] W.-K. Jung, H. Kim, Y.-C. Park, J.-W. Lee, S.-H. Ahn, Smart sewing work measurement system using IoT-based power monitoring device and approximation algorithm, *International Journal of Production Research* (2019) 1–15.

[57] H.-S. Yoon, J.-Y. Lee, M.-S. Kim, E. Kim, Y.-J. Shin, S.-Y. Kim, S. Min, S.-H. Ahn, Power Consumption Assessment of Machine Tool Feed Drive Units, *International Journal of Precision Engineering and Manufacturing–Green Technology* 7(2) (2020) 455–464.

[58] T. Benkedjough, K. Medjaher, N. Zerhouni, S. Rechak, Health assessment and life prediction of cutting tools based on support vector regression, *Journal of Intelligent Manufacturing* 26(2) (2015) 213–223.

[59] X. Li, W. Zhang, Q. Ding, J.-Q. Sun, Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation, *Journal of Intelligent Manufacturing* 31(2) (2020) 433–452.

[60] G.F. Bin, J.J. Gao, X.J. Li, B.S. Dhillon, Early fault diagnosis of rotating machinery based on wavelet packets–Empirical mode decomposition feature extraction and neural network, *Mechanical Systems and Signal Processing* 27 (2012) 696–711.

[61] Z. Tian, An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring, *Journal of Intelligent Manufacturing* 23(2) (2012) 227–237.

[62] X. Zhang, Y. Liang, J. Zhou, Y. zang, A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM, *Measurement* 69 (2015) 164–179.

[63] B.-S. Yang, X. Di, T. Han, Random forests classifier for machine fault diagnosis, *Journal of Mechanical Science and Technology* 22(9) (2008) 1716–1725.

[64] X. Li, Q. Ding, J.-Q. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, *Reliability Engineering & System Safety* 172 (2018) 1–11.

[65] C. Wang, A. Santoso, S. Mathulaprangsan, C. Chiang, C. Wu, J. Wang, Recognition and retrieval of sound events using sparse coding convolutional neural network, 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 589–594.

[66] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86(11) (1998) 2278–2324.

[67] R. Ziani, A. Felkaoui, R. Zegadi, Bearing fault diagnosis using multiclass support vector machines with binary particle swarm

optimization and regularized Fisher's criterion, *Journal of Intelligent Manufacturing* 28(2) (2017) 405–417.

[68] H. Li, X. Lian, C. Guo, P. Zhao, Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination, *Journal of Intelligent Manufacturing* 26(1) (2015) 189–198.

[69] A. Ragab, M.–S. Ouali, S. Yacout, H. Osman, Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan-Meier estimation, *Journal of Intelligent Manufacturing* 27(5) (2016) 943–958.

[70] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, Y. Kawaguchi, Anomalous Sound Detection Based on Interpolation Deep Neural Network, *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 271–275.

[71] C.–S.A. Gong, H.–C. Lee, Y.–C. Chuang, T.–H. Li, C.–H.S. Su, L.–H. Huang, C.–W. Hsu, Y.–S. Hwang, J.–D. Lee, C.–H. Chang, Design and Implementation of Acoustic Sensing System for Online Early Fault Detection in Industrial Fans, *Journal of Sensors* 2018 (2018) 4105208.

[72] B. Bayram, T.B. Duman, G. Ince, Real time detection of acoustic anomalies in industrial processes using sequential autoencoders, *Expert Systems* 38(1) (2021) e12564.

[73] H. Yun, H. Kim, E. Kim, M.B.G. Jun, Development of internal sound sensor using stethoscope and its applications for machine monitoring, *Procedia Manufacturing* 48 (2020) 1072–1078.

[74] A. Alzahrani, R. Liu, J. Kolodziej, Acoustic Assessment of an End Mill for Analysis of Tool Wear, the Annual Conference of the PHM Society, Philadelphia, Pennsylvania, USA, 2018.

[75] K. Wasmer, T. Le–Quang, B. Meylan, S.A. Shevchik, In Situ Quality Monitoring in AM Using Acoustic Emission: A Reinforcement Learning Approach, *Journal of Materials Engineering and Performance* 28(2) (2019) 666–672.

[76] F. Klocke, B. Döbbeler, T. Pullen, T. Bergs, Acoustic emission signal source separation for a flank wear estimation of drilling tools, *Procedia CIRP* 79 (2019) 57–62.

[77] P. Heilmann, R. Weiss, R. Weigel, L. Schwarz, Emission monitoring of machines using equally distributed wireless acoustic sensor nodes, *2017 IEEE SENSORS*, 2017, pp. 1–3.

[78] Z. Li, H. Zhang, D. Tan, X. Chen, H. Lei, A novel acoustic emission detection module for leakage recognition in a gas pipeline valve, *Process Safety and Environmental Protection* 105 (2017) 32–40.

- [79] Y. Alsouda, S. Pllana, A. Kurti, IoT-based Urban Noise Identification Using Machine Learning: Performance of SVM, KNN, Bagging, and Random Forest, 2019.
- [80] H. Marihart, G. Lackner, G. Schauritsch, Structural health monitoring using acoustic emission on metallic components in industrial plants, 33rd European Conference on Acoustic Emission Testing, Senlis, France, September 12–14, 2018 (EWGAE 2018) (2018).
- [81] H. Kulhandjian, N. Ramachandran, M. Kulhandjian, C. D'Amours, Human Activity Classification in Underwater using Sonar and Deep Learning, International Conference on Underwater Networks & Systems October 2019, 2019, pp. 1–5.
- [82] L. Nanni, Y.M.G. Costa, R.L. Aguiar, C.N. Silla, S. Brahnam, Ensemble of deep learning, visual and acoustic features for music genre classification, *Journal of New Music Research* 47(4) (2018) 383–397.
- [83] W. Zhang, Machine Learning Approaches to Predicting Company Bankruptcy, *Journal of Financial Risk Management* 6(4) (2017) 364–374.
- [84] E. Sejdić, I. Djurović, J. Jiang, Time-frequency feature representation using energy concentration: An overview of recent advances, *Digital Signal Processing* 19(1) (2009) 153–183.
- [85] <https://dev.to/trekhleb/playing-with-discrete-fourier-transform-algorithm-in-javascript-53n5>.
- [86] H. Meng, T. Yan, F. Yuan, H. Wei, Speech Emotion Recognition From 3D Log-Mel Spectrograms With Deep Learning Network, *IEEE Access* 7 (2019) 125868–125881.

## 초 록

스마트공장은 4차 산업혁명을 주제로 한 제조공정 분야의 주요 키워드다. 스마트 팩토리를 실현하기 위해서는 모든 기기를 중앙집중식 시스템과 연결된 스마트 기기로 만들어 실시간으로 정보를 교환할 수 있도록 하는 것이 필수적이다. 소리는 다양한 장치의 상태 정보를 동시에 담을 수 있고, 마이크만 사용하여 기기 외부에서 쉽게 녹음할 수 있기 때문에 기기를 스마트 기기로 만드는 효율적인 수단이 될 수 있다. 본 연구에서는 소리 분석을 통한 멀티 디바이스 작동 모니터링 시스템을 개발하였다. 소리 획득을 위한 마이크를 기기 외부에 설치하고 여러 기기의 소리를 동시에 녹음했다. 로그멜스펙트로그램과 합성곱 신경망(CNN)으로 녹음된 소리를 분석해 71~92%의 정확도로 3개 장치의 작동 상태를 탐지할 수 있었다. 성능 향상을 위해 강도가 다른 개별 기기 작동 사운드의 구성을 통해 가상 데이터 세트를 생성하여 학습시켰으며, 이를 통해 정확도를 87%~99%까지 높일 수 있으며, 필요한 사운드 데이터 양을 줄일 수 있다. 개발된 시스템은 수작업 장치를 사용한 작업장 및 CNC 기계가 설치된 공장 환경, 알루미늄 주조공장 등에 적용되어 성공적으로 모니터링을 수행하였다.

**주요어** : 다장비 모니터링, 음성 기반 모니터링, 합성곱 신경망 (CNN), 스마트 팩토리

**학 번** : 2016-30178