Anomalous sound detection for industrial machine with supervised and unsupervised learning algorithms: Draft

Min-Gyu Kim Adhika Retnanto Ryan Roby Sophia Shi

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Introduction and Background

Industrial machinery can break down for various reasons: contamination, environmental strains, human errors, device degradation, etc. The malfunctioning of a single piece of machinery can affect the equipment and operations upstream and/or downstream, and potentially impose devastating effects on the entire industrial process. Detecting machine malfunctions in a timely and accurate manner is essential to avoid an excessive amount of operation downtime, resource waste, productivity reduction, maintenance costs, long-term damages, and most importantly, safety concerns. Active research has been developing methods in response to the growing demand for malfunctioning industrial machinery investigation and inspection (MIMII). Several possible decision characteristics of machinery have been considered, e.g. pressure, temperature, vibration, and acoustics. The latter approach, anomalous sound detection (ASD), is detecting anomalies through sound. The sound frequency spectrum behaves differently for normal and abnormal operations, in noisy environments, as well as in different malfunctioning events. Sound detection is a technique that holds the potential for low costs and high efficiency and will be the focus of our project.

There are a few challenges to be addressed regarding MIMII using acoustic detection. (1) While sound data under normal operations can be easily obtained, available data for the wide range of malfunctioning events is limited, and there are unprecedented events that we cannot obtain data from (Purohit et al., 2019; Tanabe et al., 2021). (2) The normal dataset is more challenging to work with for non-stationary machinery, where the statistical properties of the sound data change with time, sometimes in a seasonal pattern (e.g. a valve might open fully or to a certain percentage depending on the demand) (Purohit et al., 2019). (3) Changes to the factory operation or to the ambient environment result in domain shifts in the dataset, meaning the data features shift to a different distribution. Normal sounds could potentially be monitored as abnormal if such shifts are not addressed in the model, thus reducing the model performance significantly (Tanabe et al., 2021). The objective of this proposed project is to produce a robust machine-learning model that detects machine malfunctioning from operational sound recordings that are masked by background noise while addressing the aforementioned challenges.

The specific purposes of this study are:

- Proposing a new supervised anomaly detection method based on the CNN model as a classifier
- Evaluating the performance feature extraction methods (STFT, MFCCs, GFCCs, and Mel-spectrogram), which convert the audio to image data
- Assessment of the accuracy and performance of the proposed anomaly detection model
- Exploring the over-fitting issue by increasing epochs

2. Data Description

Audio data from the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset made available by Hitachi, Ltd. will be used for the purposes of our project (Purohit et al., 2019). The dataset consists of sound samples that are generated from four types of industrial machines, including valves, pumps, fans, and slide rails. The dataset was collected using a circular microphone manufactured by System in Frontier Inc., TAMAGO-03. The microphone is a circular array that consists of eight distinct microphones with a sampling rate of 16kHz and 16 bits per sample, while also enabling single- and multi-channel-based approaches. The microphone was placed 50 cm from the pumps, fans, and slide rails and 10 cm from the valves. 10-second sound intervals were collected.

Using the measurements, the authors recorded seven sound files for four individual product models for each of the industrial machines, of which four are publicly available.

Table 1. MIMMI dataset content details (Purohit et al., 2019)

Machine type	Model ID	Normal samples samples	Abnormal samples
Valve	00	991	119
	02	708	120
	04	1,000	120
	06	992	120
Pump	00	1,006	143
	02	1005	111
	04	702	100
	06	1,036	102
Fan	00	1,011	407
	02	1,016	359
	04	1,033	348
	06	1,015	361
Slide rail	00	1,068	356
	02	1,068	267
	04	534	178
	06	534	89
Total		14,719	3,300

These sound files reflect realistic scenarios in a typical factory. More specifically, within the individual models, the data in each model consists of up to 10,000 seconds of normal sound and 1,000 seconds of abnormal sound. Anomalous sounds include contamination, leakage, rotating unbalance, etc. In addition, the normal and abnormal sounds were mixed with background noise from various factories. In total, 26,092 normal and 6,065 abnormal sounds were recorded. Currently, because the MIMII dataset website provides only the model ID 00, 02, 04, and 06, we use these data for this project, which are 14,719 normal and 3,300 abnormal sounds. Table 1 shows the details of the dataset.

3. Methods

In this section, we provide the methodologies of this study, which contain data pre-processing and preparation (Sec. 3.1.1), feature extraction methods (Sec. 3.1.2), proposed model architecture (Sec. 3.1.3), results visualization (Sec. 3.1.4), and evaluation metrics (Sec. 3.1.5).

Figure 1 shows the overall supervised machine learning procedure in this study.

3.1. Supervised learning approach

3.1.1. Data pre-processing and preparation

Since the MIMII dataset we apply in this study is well-organized and cleaned, we do not need additional pre-processing. However, the datasets have imbalanced numbers of normal and abnormal samples in each dataset, i.e. much larger normal cases. To avoid the inaccurate AUC scores caused by the imbalanced dataset, we randomly select the same numbers of normal samples with the abnormal samples for each dataset. Then, we divide the dataset into train and test datasets with a 0.5:0.5 ratio. Lastly, we label the normal cases to 0 and the abnormal cases to 1.

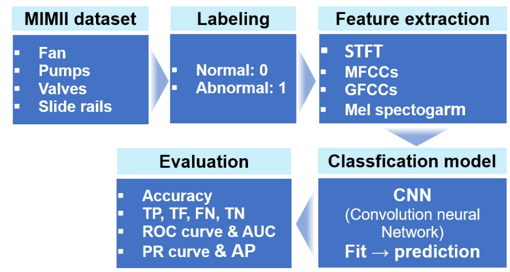

Figure 1. Overall supervised machine learning procedure

Figure 2 shows the data preparation process for supervised learning.

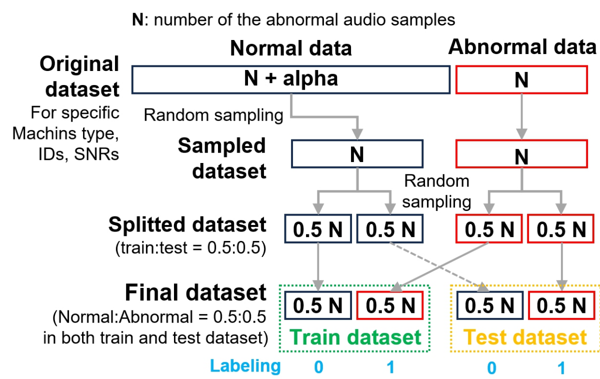

Figure 2. Data preparation for the supervised learning

3.1.2. Feature extraction methods

To implement the machine learning algorithm using the sound data, we need an appropriate feature extraction method. Because the CNN model requires an image format as input data, we convert the audio data (i.e. wav file) from the MIMII dataset to image files using various feature extraction methods, i.e. STFT (Short-term Fourier Transform), MFCCs (Mel-frequency cepstral coefficients), GFCCs (Gammatone frequency cepstral coefficients), and Mel-spectrogram.

Each feature extraction method requires several hyperparameters. The hyperparameters we apply in this study are as follows.

- STFT: nperseg = 1024, noverlap = 512
- MFCC: n_mfcc = 200
- GFCC: nfilts =250, num_ceps=250
- Mel-spectrogram: n_fft=1024, hop_length=512, n_mels=128, power=2

Where, nperseg: number of data points used in each STFT block, noverlap: number of points of overlap between blocks, n_mfcc: number of MFCCs to return, nfilts: number of filter banks, num_ceps: number of cepstral coefficients to return, n_fft: number of points for the FFT, hop_length: number of samples between successive frames, n_mels: number of Mel bands to generate, power: exponent for the magnitude spectrogram.

The above hyperparameters can affect the performance of the classification model. Thus, they should be selected carefully. However, because optimizing the hyperparameters needs many trial-and-error processes, we just choose some heuristic values, not too large and not too small.

Figure 3 shows the examples of the feature extraction results. In some cases, we can easily recognize the abnormal images visually.

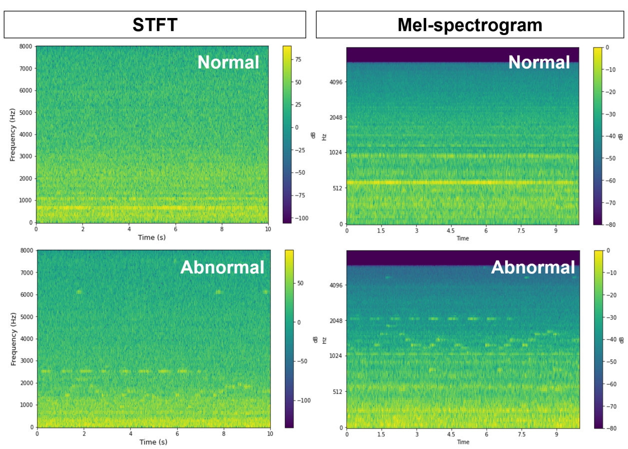

Figure 3. Visualization examples of the feature extraction results (machine: fan, STFT and Mel-spectrogram methods)

3.1.3. Proposed model architecture

We propose a new classification method based on various feature extraction methods and the CNN model. A Convolutional Neural Network (CNN) is a type of deep learning algorithm designed for processing and analyzing visual data, such as images and videos. The CNN models are particularly effective in tasks like image recognition, object detection, and classification. There are several attempts to implement the CNN model for anomaly sound detection. (Morita et al., 2021; Zhao, 2020) In our study, the CNN model is implemented as a classifier based on spectrograms extracted from the audio sound dataset as an input.

Table 2 and Figure 4 show the proposed CNN model architecture using four different feature extraction methods. The output shapes at each layer have different sizes, which depend on the hyperparameters of the feature extraction methods.

For the model fitting, we apply the *binary cross entropy* loss function, *Adam* optimizer, 0.001 learning rate, 16 batch size, and 8 different epochs between 5-40.

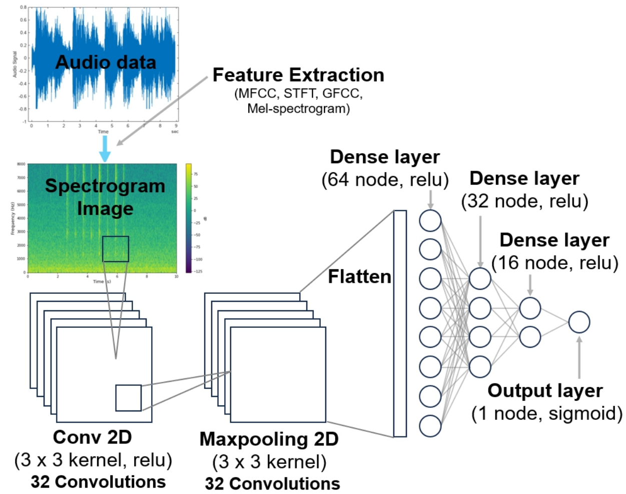

Figure 4. Conceptual image of the *spectrogram + CNN architecture* in this study for the supervised learning

3.1.4. Visualization methods of the results

To evaluate the performance of individual datasets, we plot the predicted score [0,1] from the output of the CNN model with the labeled true status 0,1 for the train and test datasets.

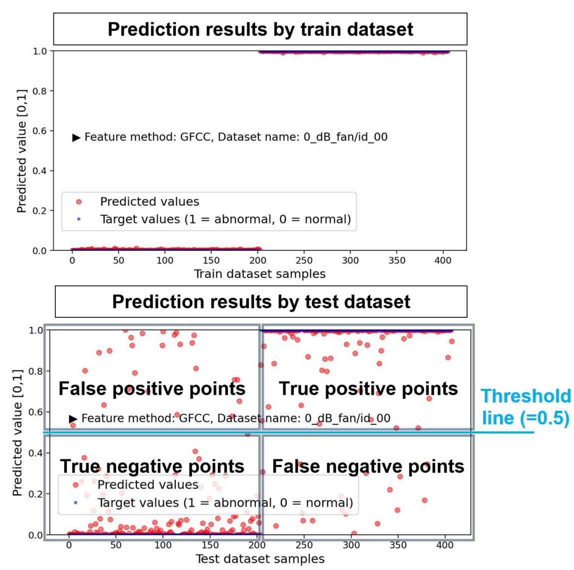

Figure 5. Visualization example of the model prediction results for each dataset

Figure 5 shows the example including the predicted values [0,1] by the CNN model (red circles) with the true values (blue small circles). In this figure, we can see the distribution of the results visually, and which threshold is appropriate for the classification. However, because the threshold can have

Table 2. Proposed CNN model architecture

Layer	Output shape STFT	MFCCs	GFCCs	Mel-spectrogram
Feature extraction methods	(514 X 314)	(128 X 431)	(1000 X 250)	(128 X 431)
Convolution 2D (3X3)	(512 X 312, 32)	(126 X 429, 32)	(998 X 248, 32)	(126 X 429, 32)
Max pooling 2D (3X3)	(170 X 104, 32)	(42 X 143, 32)	(332 X 82, 32)	(42 X 143, 32)
Flatten	565,760	192,192	871,168	192,192
Dense ('relu' activation)			64	
Dense ('relu' activation)			32	
Dense ('relu' activation)			16	
Dense ('sigmoid' activation)			1	

very different values according to the model IDs and SNRs (signal-to-noise ratios), to avoid over-fitted classification, we assume the threshold is 0.5 for all datasets. In this study, we made a total of 6,144 figures considering four feature extraction methods, four machine types, four model IDs, three SNRs, and 8 different epochs.

3.1.5. Evaluation metrics

In this study, we basically develop a binary classification model. Therefore, for evaluation, the ROC curve with AUC score and PR curve with AP score would be reasonable choices. In addition, we also assess other results such as accuracy and confusion matrix.

Figure 6 shows the example of the ROC and PR curve for specific datasets in this study.

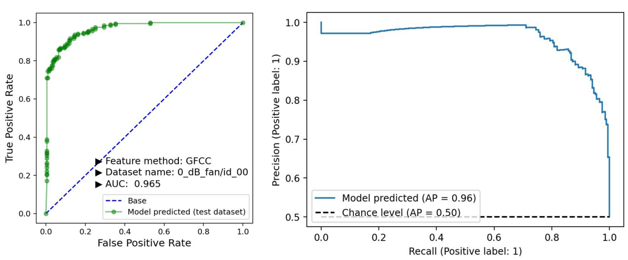

Figure 6. Visualization example of the model prediction results for each dataset

We also use the confusion matrix to show the result visually and compare the model result between different datasets.

4. Results and discussion

4.1. Result of the supervised learning model

4.1.1. Comparison of AUC scores between the MIMII baseline model and proposed method

In the previous research (Purohit et al., 2019), the author provides the baseline model based on the unsupervised auto-encoder model. In this study, we compare our results to their model using the AUC scores. According to the results, the proposed supervised CNN model outperform s the MIMII's (Purohit et al., 2019) unsupervised model with any feature extraction methods considered in this study.

Figure 7 shows the average AUC scores according to the four machine types and four feature extraction methods by different epochs. Generally, increasing epochs can help to improve the model performance before occurring the over-fitting issues. In this result, when we apply the 25 epochs, the AUC scores are improved by 0.09-0.15 for the valve, 0.01-0.09 for the pump, 0.01-0.06 for the Fan, and 0.03-0.08 for the slide rail compared to the 5 epochs cases.

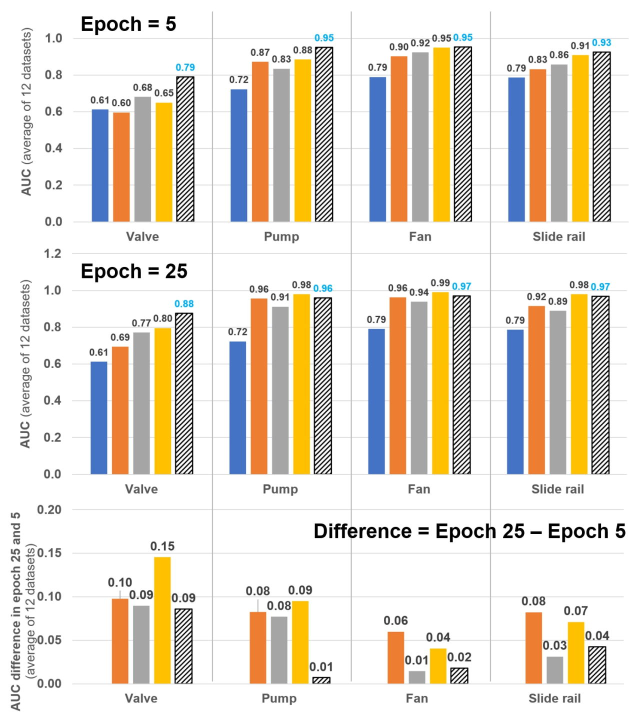

Figure 7. Comparison of the AUC scores between MIMII baseline model and proposed CNN model by different feature extraction methods (STFT, MFCCs, GFCCs, and Mel-spectrogram)

4.1.2. Impact of the SNRs in the model performance

According to the confusion matrix, the SNRs (Signal-to-noise ratios) show significant impacts on the model performance. With the higher SNR, i.e. 6 dB, the model performances also show higher accuracy, while the model performances worsen with the lower SNR, i.e. -6 dB. Because this result indicates that getting the improved model performance with the lower SNR is difficult, further research is required to overcome this limitation.

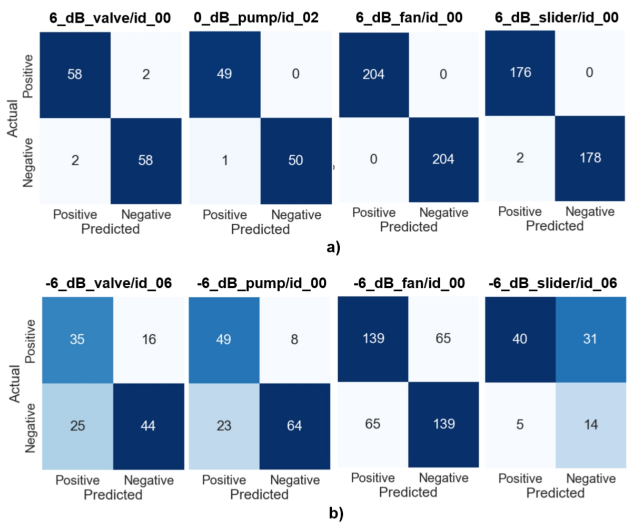

Figure 8. Confusion matrices for the best and worst cases for four machine type with the STFT method at epoch 25: a) best results, b) Worst results

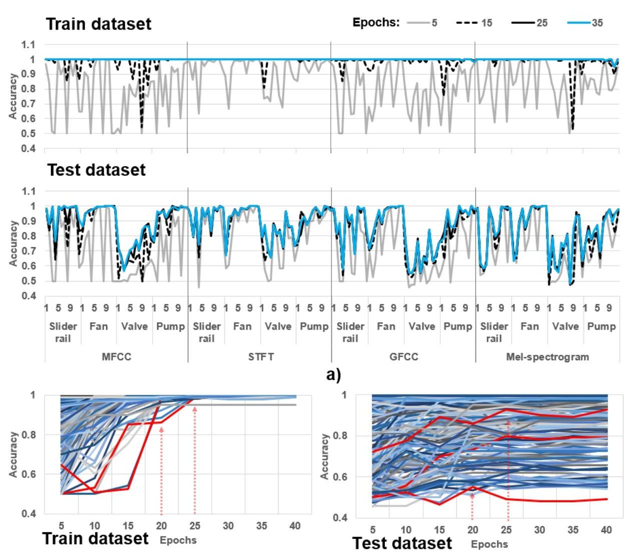

Figure 9. Accuracy changes of 192 individual models by increasing learning epochs

Figure 8 shows the confusion matrix including the a) best result and b) worst result for four machine types.

4.1.3. Assessment for the over-fitting by increasing epochs

We observe that there are no significant over-fitting problems by increasing the epochs because the accuracy for each model tends to converge certain values except in several cases. For instance, in Figure 9 b), red lines show the over-fitting tendency in larger epochs. Based on these results, we conclude that 25-30 epochs may give the best results for our model. In other words, we need to stop the learning process around 25-30 epochs.

Figure 9 shows the train and test accuracy of 192 cases that represent machine types, domains, SNR level, and feature extraction methods along different Epochs.

4.2. Limitations of the supervised learning method

The supervised learning approach has several limitations such as:

- It can not adapt to the domain shift. In other words, it might be not a robust method in realistic industrial environments.
- It struggles with the generalization under the various condition changes such as different SNRs and machine models.
- Obtaining and labeling the audio samples from all possible abnormal cases for supervised learning is almost impossible to work.

4.3. Discussion about the unsupervised learning method

4.3.1. Unsupervised learning

(Min-Gyu's opinion) If we could not get any result from the unsupervised learning, we can discuss it here including the necessity of unsupervised learning, research cases in literature, possible approaches, etc.

5. Conclusions

In this study, we propose a new anomaly detection methods.

(Min-Gyu's opinion) We need to summarize the major contents of this report.

(Min-Gyu's opinion) The length of the report is limited up to 6 page except *References*. I think we need to add 1 more page, or reduce this draft to add more new contents.

6. Future works

We propose the future works to improve our results such as:

- Application of the transfer learning methods for generalization to the various domain drift
- Incorporation of the ensemble algorithms as a classifier based on the CNN model's flatten output data.
- Investigation about the unsupervised learning methods
- Exploring the other image classification algorithms

Accessibility

The Python code and relevant documents are provided in this project link: https://github.com/minsky97/ECE381K_AML_term_project.

References

Morita, K., Yano, T., and Tran, K. Anomalous sound detection using cnn-based features by self supervised learning. *Tech. Rep., DCASE2021 Challenge*, 2021.

Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., and Kawaguchi, Y. MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. 2019. doi: 10.48550/ARXIV.1909.09347.

Tanabe, R., Purohit, H., Dohi, K., Endo, T., Nikaido, Y., Nakamura, T., and Kawaguchi, Y. MIMII Due: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection with Domain Shifts Due to Changes in Operational and Environmental Conditions. In *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 21–25, New Paltz, NY, USA, October 2021. IEEE. doi: 10.1109/WASPAA52581.2021.9632802.

Zhao, J. Anomalous sound detection based on convolutional neural network and mixed features. In *Journal of Physics: Conference Series*, volume 1621, pp. 012025. IOP Publishing, 2020.

A. All AUC scores of the individual dataset under the supervised learning with epoch 5-40

This is just a test for the appendix. I am not sure attaching these tables be meaningful. The length of appendix is limited **up to 4 pages**. So, if we have better data for appendix, I think we need to change these tables.

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	00	0.68	0.55	0.62	0.80	0.62	0.44	0.86	0.66	0.65	0.98	0.93	0.81	0.88	0.56	0.47
	02	0.66	0.59	0.57	0.75	0.62	0.61	0.58	0.54	0.50	0.91	0.86	0.76	0.98	0.72	0.50
	04	0.64	0.65	0.50	0.76	0.71	0.57	0.64	0.60	0.52	0.87	0.77	0.56	0.96	0.87	0.59
	06	0.70	0.66	0.53	0.70	0.64	0.57	0.58	0.53	0.49	0.78	0.62	0.62	0.56	0.52	0.56
	Avg.	**0.67**	**0.61**	**0.56**	**0.75**	**0.65**	**0.55**	**0.66**	**0.58**	**0.54**	**0.88**	**0.80**	**0.69**	**0.84**	**0.67**	**0.53**
Pump	00	0.84	0.65	0.58	0.98	0.88	0.78	0.92	0.87	0.81	0.96	0.91	0.82	0.94	0.85	0.82
	02	0.45	0.46	0.52	1.00	0.85	0.30	0.78	0.69	0.73	1.00	0.96	0.98	0.67	0.58	0.74
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.97	1.00	1.00	0.98	0.96	0.94	0.72
	06	0.94	0.76	0.61	1.00	0.98	0.86	1.00	0.93	0.79	1.00	0.97	0.85	0.99	0.98	0.84
	Avg.	**0.81**	**0.71**	**0.66**	**0.99**	**0.93**	**0.73**	**0.92**	**0.87**	**0.82**	**0.99**	**0.96**	**0.91**	**0.89**	**0.84**	**0.78**
Fan	00	0.75	0.63	0.57	1.00	0.99	0.82	0.90	0.74	0.73	1.00	0.95	0.70	0.99	0.82	0.71
	02	0.99	0.83	0.68	1.00	1.00	0.96	0.99	0.93	0.79	1.00	1.00	0.90	1.00	0.94	0.72
	04	0.92	0.75	0.57	1.00	1.00	0.90	1.00	0.99	0.79	1.00	1.00	0.89	1.00	1.00	0.92
	06	0.99	0.97	0.83	1.00	0.93	0.78	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**0.98**	**0.87**	**0.97**	**0.91**	**0.82**	**1.00**	**0.99**	**0.87**	**1.00**	**0.94**	**0.83**
Slide rail	00	0.99	0.99	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.98	1.00	1.00	1.00
	02	0.93	0.79	0.74	0.97	0.95	0.87	0.96	0.92	0.85	1.00	0.95	0.87	1.00	0.86	0.57
	04	0.88	0.78	0.61	1.00	0.96	0.94	0.97	0.85	0.83	1.00	1.00	0.99	0.96	0.75	0.61
	06	0.71	0.56	0.52	0.92	0.77	0.52	0.60	0.63	0.40	0.94	0.95	0.44	1.00	0.94	0.63
	Avg.	**0.88**	**0.78**	**0.70**	**0.97**	**0.92**	**0.83**	**0.88**	**0.85**	**0.76**	**0.98**	**0.97**	**0.82**	**0.99**	**0.88**	**0.70**

Figure A1. AUC scores for the individual test dataset with the epoch = 5

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	00	0.68	0.55	0.62	0.83	0.70	0.65	0.84	0.78	0.66	1.00	0.94	0.88	0.86	0.72	0.54
	02	0.66	0.59	0.57	0.88	0.63	0.66	0.58	0.54	0.53	0.91	0.86	0.67	0.99	0.77	0.72
	04	0.64	0.65	0.50	0.84	0.83	0.60	0.66	0.68	0.54	0.90	0.86	0.76	0.96	0.88	0.61
	06	0.70	0.66	0.53	0.83	0.69	0.61	0.77	0.56	0.58	0.83	0.77	0.68	0.58	0.52	0.54
	Avg.	**0.67**	**0.61**	**0.56**	**0.85**	**0.71**	**0.63**	**0.71**	**0.64**	**0.58**	**0.91**	**0.86**	**0.75**	**0.85**	**0.72**	**0.61**
Pump	00	0.84	0.65	0.58	0.99	0.95	0.80	0.96	0.92	0.82	0.98	0.92	0.80	0.98	0.88	0.82
	02	0.45	0.46	0.52	1.00	0.93	0.88	0.98	0.89	0.78	0.99	1.00	0.95	0.93	0.79	0.87
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.97	1.00	1.00	0.96	0.96	0.94	0.79
	06	0.94	0.76	0.61	1.00	0.99	0.98	1.00	0.98	0.89	1.00	0.98	0.91	0.99	0.98	0.89
	Avg.	**0.81**	**0.71**	**0.66**	**1.00**	**0.96**	**0.91**	**0.99**	**0.95**	**0.87**	**0.99**	**0.98**	**0.90**	**0.97**	**0.90**	**0.84**
Fan	00	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.88	0.74	1.00	0.96	0.74	1.00	0.91	0.68
	02	0.99	0.83	0.68	1.00	1.00	0.97	1.00	0.97	0.85	1.00	1.00	0.93	1.00	0.96	0.72
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.91	1.00	1.00	0.96	1.00	1.00	0.93
	06	0.99	0.97	0.83	1.00	1.00	0.90	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**1.00**	**0.94**	**1.00**	**0.96**	**0.87**	**1.00**	**0.99**	**0.91**	**1.00**	**0.97**	**0.83**
Slide rail	00	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	0.99	0.93	0.99	0.95	0.90	1.00	0.97	0.89	1.00	0.94	0.61
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.95	0.89	1.00	1.00	1.00	0.99	0.79	0.57
	06	0.71	0.56	0.52	1.00	0.88	0.80	0.80	0.64	0.49	1.00	0.99	0.82	1.00	0.92	0.69
	Avg.	**0.88**	**0.78**	**0.70**	**1.00**	**0.97**	**0.93**	**0.95**	**0.89**	**0.82**	**1.00**	**0.99**	**0.92**	**1.00**	**0.91**	**0.72**

Figure A2. AUC scores for the individual test dataset with the epoch = 10

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	**0 dB**	**-6 dB**	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	**00**	0.68	0.55	0.62	0.82	0.75	0.71	0.87	0.78	0.71	1.00	0.93	0.89	0.92	0.79	0.50
	02	0.66	0.59	0.57	0.91	0.66	0.72	0.59	0.54	0.53	0.93	0.86	0.73	1.00	0.79	0.79
	04	0.64	0.65	0.50	0.91	0.85	0.61	0.72	0.69	0.55	0.91	0.91	0.78	0.97	0.89	0.62
	06	0.70	0.66	0.53	0.87	0.73	0.68	0.85	0.67	0.62	0.90	0.77	0.71	0.62	0.50	0.64
	Avg.	**0.67**	**0.61**	**0.56**	**0.88**	**0.75**	**0.68**	**0.76**	**0.67**	**0.60**	**0.93**	**0.87**	**0.78**	**0.87**	**0.74**	**0.64**
Pump	**00**	0.84	0.65	0.58	1.00	0.97	0.82	0.98	0.91	0.81	0.98	0.93	0.80	0.99	0.92	0.80
	02	0.45	0.46	0.52	1.00	0.98	0.95	0.99	0.96	0.85	1.00	1.00	0.96	0.95	0.80	0.90
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.96	0.94	0.79
	06	0.94	0.76	0.61	1.00	0.99	0.99	1.00	0.98	0.89	1.00	0.98	0.94	0.99	0.98	0.91
	Avg.	**0.81**	**0.71**	**0.66**	**1.00**	**0.99**	**0.94**	**0.99**	**0.96**	**0.88**	**1.00**	**0.98**	**0.92**	**0.97**	**0.91**	**0.85**
Fan	**00**	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.94	0.75	1.00	0.98	0.75	1.00	0.92	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.87	1.00	1.00	0.95	1.00	0.96	0.72
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.94	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**1.00**	**0.97**	**1.00**	**0.98**	**0.89**	**1.00**	**0.99**	**0.91**	**1.00**	**0.97**	**0.84**
Slide rail	**00**	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.94	0.99	0.96	0.90	1.00	0.98	0.90	1.00	0.95	0.62
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.96	0.94	1.00	1.00	0.99	1.00	0.82	0.60
	06	0.71	0.56	0.52	1.00	0.95	0.80	0.84	0.73	0.54	1.00	0.96	0.75	1.00	0.92	0.69
	Avg.	**0.88**	**0.78**	**0.70**	**1.00**	**0.99**	**0.93**	**0.96**	**0.91**	**0.84**	**1.00**	**0.98**	**0.91**	**1.00**	**0.92**	**0.73**

Figure A3. AUC scores for the individual test dataset with the epoch = 15

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	**0 dB**	**-6 dB**	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	**00**	0.68	0.55	0.62	0.85	0.78	0.72	0.87	0.79	0.71	1.00	0.93	0.89	0.91	0.83	0.58
	02	0.66	0.59	0.57	0.92	0.71	0.74	0.61	0.55	0.54	0.92	0.86	0.90	0.99	0.78	0.81
	04	0.64	0.65	0.50	0.93	0.84	0.63	0.73	0.69	0.57	0.91	0.91	0.78	0.97	0.91	0.62
	06	0.70	0.66	0.53	0.88	0.75	0.67	0.86	0.69	0.65	0.90	0.77	0.70	0.64	0.52	0.65
	Avg.	**0.67**	**0.61**	**0.56**	**0.89**	**0.77**	**0.69**	**0.77**	**0.68**	**0.62**	**0.93**	**0.87**	**0.82**	**0.88**	**0.76**	**0.66**
Pump	**00**	0.84	0.65	0.58	1.00	0.97	0.83	0.99	0.91	0.81	0.98	0.93	0.81	0.99	0.90	0.77
	02	0.45	0.46	0.52	1.00	1.00	0.98	0.99	0.98	0.88	1.00	1.00	1.00	0.96	0.80	0.90
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.95	0.94	0.82
	06	0.94	0.76	0.61	1.00	0.99	0.99	0.99	0.99	0.91	1.00	1.00	0.94	1.00	0.98	0.89
	Avg.	**0.81**	**0.71**	**0.66**	**1.00**	**0.99**	**0.95**	**0.99**	**0.97**	**0.90**	**1.00**	**0.98**	**0.93**	**0.98**	**0.91**	**0.85**
Fan	**00**	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.96	0.77	1.00	0.98	0.75	1.00	0.93	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.86	1.00	1.00	0.95	1.00	0.97	0.73
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**1.00**	**0.97**	**1.00**	**0.99**	**0.89**	**1.00**	**0.99**	**0.92**	**1.00**	**0.97**	**0.84**
Slide rail	**00**	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.95	1.00	0.97	0.90	1.00	0.98	0.90	1.00	0.95	0.63
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.97	0.93	1.00	1.00	0.99	1.00	0.84	0.60
	06	0.71	0.56	0.52	1.00	0.96	0.81	0.90	0.73	0.56	1.00	0.96	0.78	1.00	0.94	0.70
	Avg.	**0.88**	**0.78**	**0.70**	**1.00**	**0.99**	**0.94**	**0.97**	**0.92**	**0.85**	**1.00**	**0.99**	**0.92**	**1.00**	**0.93**	**0.73**

Figure A4. AUC scores for the individual test dataset with the epoch = 20

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	00	0.68	0.55	0.62	0.88	0.78	0.71	0.87	0.79	0.71	1.00	0.93	0.89	0.93	0.84	0.59
	02	0.66	0.59	0.57	0.92	0.77	0.75	0.61	0.57	0.55	0.92	0.86	0.96	0.99	0.78	0.82
	04	0.64	0.65	0.50	0.94	0.85	0.63	0.73	0.69	0.58	0.91	0.90	0.78	0.96	0.91	0.62
	06	0.70	0.66	0.53	0.89	0.76	0.66	0.85	0.71	0.66	0.90	0.77	0.70	0.68	0.52	0.63
	Avg.	0.67	0.61	0.56	0.91	0.79	0.69	0.77	0.69	0.63	0.93	0.87	0.83	0.89	0.76	0.66
Pump	00	0.84	0.65	0.58	1.00	0.97	0.83	0.99	0.91	0.82	0.98	0.93	0.81	0.99	0.89	0.77
	02	0.45	0.46	0.52	1.00	1.00	0.99	0.99	0.98	0.90	1.00	1.00	0.89	0.97	0.81	0.90
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.96	0.95	0.82
	06	0.94	0.76	0.61	1.00	0.99	0.98	1.00	0.99	0.90	1.00	1.00	0.94	1.00	0.98	0.90
	Avg.	0.81	0.71	0.66	1.00	0.99	0.95	1.00	0.97	0.90	1.00	0.98	0.90	0.98	0.91	0.85
Fan	00	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.96	0.78	1.00	0.98	0.75	1.00	0.92	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.88	1.00	1.00	0.96	1.00	0.97	0.73
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	0.91	0.80	0.66	1.00	1.00	0.97	1.00	0.99	0.90	1.00	0.99	0.92	1.00	0.97	0.84
Slide rail	00	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.95	1.00	0.97	0.91	1.00	0.98	0.90	1.00	0.95	0.63
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.97	0.94	1.00	1.00	0.99	1.00	0.84	0.61
	06	0.71	0.56	0.52	1.00	0.99	0.82	0.90	0.74	0.56	1.00	0.97	0.79	1.00	0.94	0.71
	Avg.	0.88	0.78	0.70	1.00	1.00	0.94	0.97	0.92	0.85	1.00	0.99	0.92	1.00	0.93	0.74

Figure A5. AUC scores for the individual test dataset with the epoch = 25

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	00	0.68	0.55	0.62	0.88	0.79	0.72	0.87	0.79	0.71	1.00	0.93	0.89	0.94	0.84	0.60
	02	0.66	0.59	0.57	0.93	0.80	0.76	0.62	0.56	0.56	0.92	0.86	0.97	0.99	0.78	0.82
	04	0.64	0.65	0.50	0.94	0.85	0.63	0.73	0.69	0.58	0.91	0.90	0.78	0.97	0.90	0.62
	06	0.70	0.66	0.53	0.89	0.76	0.68	0.85	0.70	0.65	0.90	0.77	0.70	0.69	0.52	0.63
	Avg.	0.67	0.61	0.56	0.91	0.80	0.70	0.77	0.68	0.63	0.93	0.86	0.84	0.90	0.76	0.67
Pump	00	0.84	0.65	0.58	1.00	0.97	0.83	0.99	0.92	0.82	0.98	0.93	0.81	0.99	0.89	0.77
	02	0.45	0.46	0.52	1.00	1.00	0.99	0.99	0.98	0.88	1.00	1.00	0.97	0.96	0.81	0.91
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.96	0.94	0.82
	06	0.94	0.76	0.61	1.00	0.99	0.98	1.00	0.99	0.91	1.00	1.00	0.94	1.00	0.97	0.90
	Avg.	0.81	0.71	0.66	1.00	0.99	0.95	1.00	0.97	0.90	1.00	0.98	0.92	0.98	0.91	0.85
Fan	00	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.96	0.78	1.00	0.98	0.75	1.00	0.92	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.88	1.00	1.00	0.96	1.00	0.97	0.73
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	0.91	0.80	0.66	1.00	1.00	0.97	1.00	0.99	0.90	1.00	0.99	0.92	1.00	0.97	0.84
Slide rail	00	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.96	1.00	0.97	0.91	1.00	0.98	0.90	1.00	0.95	0.63
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.97	0.94	1.00	1.00	0.99	1.00	0.85	0.62
	06	0.71	0.56	0.52	1.00	0.99	0.83	0.91	0.75	0.56	1.00	0.97	0.80	1.00	0.93	0.71
	Avg.	0.88	0.78	0.70	1.00	1.00	0.95	0.98	0.92	0.85	1.00	0.99	0.92	1.00	0.93	0.74

Figure A6. AUC scores for the individual test dataset with the epoch = 30

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	**0 dB**	**-6 dB**	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	**00**	0.68	0.55	0.62	0.88	0.79	0.72	0.87	0.79	0.72	1.00	0.93	0.90	0.95	0.84	0.60
	02	0.66	0.59	0.57	0.93	0.81	0.76	0.62	0.56	0.55	0.92	0.86	0.97	0.99	0.78	0.82
	04	0.64	0.65	0.50	0.94	0.85	0.63	0.73	0.69	0.59	0.91	0.90	0.78	0.97	0.90	0.62
	06	0.70	0.66	0.53	0.89	0.76	0.68	0.85	0.69	0.66	0.90	0.77	0.70	0.69	0.52	0.63
	Avg.	**0.67**	**0.61**	**0.56**	**0.91**	**0.80**	**0.70**	**0.77**	**0.68**	**0.63**	**0.93**	**0.86**	**0.84**	**0.90**	**0.76**	**0.67**
Pump	**00**	0.84	0.65	0.58	1.00	0.98	0.83	0.99	0.92	0.82	0.98	0.93	0.81	0.99	0.89	0.77
	02	0.45	0.46	0.52	1.00	1.00	0.99	0.99	0.98	0.88	1.00	1.00	0.91	0.96	0.82	0.91
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.97	0.95	0.82
	06	0.94	0.76	0.61	1.00	0.99	0.98	0.99	0.99	0.91	1.00	1.00	0.94	1.00	0.97	0.90
	Avg.	**0.81**	**0.71**	**0.66**	**1.00**	**0.99**	**0.95**	**0.99**	**0.97**	**0.90**	**1.00**	**0.98**	**0.91**	**0.98**	**0.91**	**0.85**
Fan	**00**	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.96	0.78	1.00	0.98	0.75	1.00	0.93	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.88	1.00	1.00	0.96	1.00	0.97	0.73
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**1.00**	**0.97**	**1.00**	**0.99**	**0.90**	**1.00**	**0.99**	**0.92**	**1.00**	**0.97**	**0.84**
Slide rail	**00**	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.96	1.00	0.97	0.91	1.00	0.98	0.90	1.00	0.95	0.63
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.97	0.94	1.00	1.00	0.99	1.00	0.85	0.62
	06	0.71	0.56	0.52	1.00	0.99	0.83	0.91	0.75	0.56	1.00	0.97	0.81	1.00	0.94	0.71
	Avg.	**0.88**	**0.78**	**0.70**	**1.00**	**1.00**	**0.95**	**0.98**	**0.92**	**0.85**	**1.00**	**0.99**	**0.92**	**1.00**	**0.94**	**0.74**

Figure A7. AUC scores for the individual test dataset with the epoch = 35

Machine type	Model ID	Previous MIMII paper			MFCC + CNN			GFCC + CNN			STFT + CNN			Mel_spectrogram + CNN		
		Input SNR			Input SNR			Input SNR			Input SNR			Input SNR		
		6 dB	**0 dB**	**-6 dB**	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB	6 dB	0 dB	-6 dB
Valve	**00**	0.68	0.55	0.62	0.89	0.79	0.72	0.87	0.79	0.72	1.00	0.93	0.89	0.95	0.84	0.60
	02	0.66	0.59	0.57	0.93	0.81	0.76	0.61	0.56	0.56	0.92	0.86	0.97	0.99	0.78	0.81
	04	0.64	0.65	0.50	0.94	0.85	0.64	0.73	0.69	0.59	0.91	0.90	0.78	0.97	0.90	0.62
	06	0.70	0.66	0.53	0.89	0.76	0.68	0.85	0.69	0.66	0.90	0.77	0.70	0.69	0.52	0.63
	Avg.	**0.67**	**0.61**	**0.56**	**0.91**	**0.80**	**0.70**	**0.77**	**0.68**	**0.63**	**0.93**	**0.86**	**0.84**	**0.90**	**0.76**	**0.67**
Pump	**00**	0.84	0.65	0.58	1.00	0.98	0.83	0.99	0.92	0.82	0.98	0.93	0.81	0.99	0.89	0.77
	02	0.45	0.46	0.52	1.00	1.00	0.99	0.99	0.98	0.87	1.00	1.00	0.95	0.95	0.81	0.91
	04	0.99	0.95	0.93	1.00	1.00	0.99	1.00	1.00	0.98	1.00	1.00	0.97	0.97	0.95	0.82
	06	0.94	0.76	0.61	1.00	0.99	0.98	1.00	0.99	0.91	1.00	1.00	0.94	1.00	0.97	0.90
	Avg.	**0.81**	**0.71**	**0.66**	**1.00**	**0.99**	**0.95**	**1.00**	**0.97**	**0.90**	**1.00**	**0.98**	**0.92**	**0.98**	**0.91**	**0.85**
Fan	**00**	0.75	0.63	0.57	1.00	1.00	0.91	1.00	0.97	0.78	1.00	0.98	0.75	1.00	0.93	0.70
	02	0.99	0.83	0.68	1.00	1.00	0.98	1.00	0.99	0.88	1.00	1.00	0.96	1.00	0.97	0.73
	04	0.92	0.75	0.57	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.97	1.00	1.00	0.94
	06	0.99	0.97	0.83	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	Avg.	**0.91**	**0.80**	**0.66**	**1.00**	**1.00**	**0.97**	**1.00**	**0.99**	**0.90**	**1.00**	**0.99**	**0.92**	**1.00**	**0.97**	**0.84**
Slide rail	**00**	0.99	0.99	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	02	0.93	0.79	0.74	1.00	1.00	0.96	1.00	0.97	0.91	1.00	0.98	0.90	1.00	0.95	0.63
	04	0.88	0.78	0.61	1.00	1.00	1.00	1.00	0.98	0.94	1.00	1.00	0.99	1.00	0.85	0.63
	06	0.71	0.56	0.52	1.00	0.99	0.83	0.91	0.75	0.56	1.00	0.97	0.81	1.00	0.94	0.71
	Avg.	**0.88**	**0.78**	**0.70**	**1.00**	**1.00**	**0.95**	**0.98**	**0.92**	**0.85**	**1.00**	**0.99**	**0.92**	**1.00**	**0.94**	**0.74**

Figure A8. AUC scores for the individual test dataset with the epoch = 40