PAPER • OPEN ACCESS

Deep learning and acoustic approach for mechanical failure detection in industrial machinery

To cite this article: Ahmad Qurthobi and Rytis Maskeliūnas 2023 *J. Phys.: Conf. Ser.* **2673** 012032

Deep learning and acoustic approach for mechanical failure detection in industrial machinery

Ahmad Qurthobi [1,2], and Rytis Maskeliūnas[1]

[1]Department of Multimedia Engineering, Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania

[2]Department of Engineering Physics, School of Electrical Engineering, Telkom University, Bandung, Indonesia

E-mail: `rytis.maskeliunas@ktu.lt`

Abstract. Research into acoustic signal-based failure detection has developed into a subject that has attracted the attention of many researchers in recent years. Acoustic signal data collection can be performed without having to interrupt or stop the operation of the machine to be inspected. Therefore, it is very beneficial for the development of nondestructive testing and predictive maintenance. In this study, a collection of pump sound recordings that are part of the Malfunctioning Industrial Machine Investigation and Inspection dataset, known as the MIMII dataset, is used as test material. Several deep learning algorithms such as long short-term memory (LSTM), gate recurrent unit (GRU), autoencoder, and convolutional neural network (CNN) were involved and compared to determine their ability to detect failures. Based on the training results with 300 epochs and a learning rate of 10^{-6}, it was found that CNN produced the classification with the highest accuracy compared to the other algorithms. In addition, the CNN algorithm is also capable of performing classification amidst the problem of imbalance in the amount of data.

1. Introduction

For every industrial machinery, a machine's state of health could be defined from its surface temperature, body vibration, or operation sound [1, 2]. An expert operator could utilize these variables to observe the presence of abnormalities. For example, a pump is assumed to be operating abnormally when sounds with higher or lower frequencies than normal are heard. In addition, the pump is also assumed to operate abnormally when there is an overly frequent startup sound. However, if pumps, or any type of industrial machinery in general, are operated in a high-noise environment, special attention needs to be paid as this can make it difficult for human labour to determine operating conditions. In addition, high noise conditions can also jeopardize the health and safety of the operators themselves [3].

One of the research topics that has been gaining attention in recent years is anomaly detection in industrial machinery [4–7]. Research utilizing acoustic signals for the classification of machine operating conditions is considered very useful in developing machine preventive maintenance, machine life prediction, and quality control processes. In addition to the development of more capable and affordable hardware, this topic has also received more attention from the academic world due to the large selection of artificial intelligence and deep learning algorithms that facilitate the classification process.

Long short-term memory (LSTM), gate repeating unit (GRU), autoencoder (AE), and convolutional neural network (CNN) are some examples of deep learning algorithms for classification. Long short-term memory (LSTM) is one of the deep learning algorithms that can be used for classification, prediction, and control [8–11]. It can learn complex patterns and relationships in the input data, making it a valuable tool for a wide range of tasks in various fields such as finance, healthcare, and natural language processing [12–14]. It also has the capacity to discern audio signals based on their anticipated properties, which is a very dependable feature for audio classification. on the other side, GRU is a deep learning algorithm that is similar to LSTM but has fewer parameters and an additional forget gate. This algorithm was coined by Cho et al [15] and works with the gate mechanism in the recurrent neural network. Apart from LSTM and CNN, which are developments of recurrent neural networks and are indeed commonly used algorithms for processing speech signals, autoencoder is also able to utilize in acoustic signals classification [16–19]. It is a type of deep learning algorithm that is capable of performing classification, filtering, and reconstruction at the same time and adequate to copy input data on the output side. In general, an autoencoder works with two main functions: the encoder part which functions to compress the input data, and the decoder part which reconstructs the data to its original form. Lastly, another option for audio classification issues is CNN [20–23]. It was first developed in 1995 by Lecun and Bengio and is very commonly used in image-based classification [24]. However, it has also proven capable of classifying audio signals by utilizing features such as mel-spectrogram, chroma, or short-time Fourier transform extracted from the signal.

This paper presents a comparison of the ability of several deep learning algorithms to classify operating conditions based on acoustic data of industrial machinery. AE, LSTM, GRU, and CNN were tested using a collection of acoustic signals. A group of pump sound files with *.wav extension which is part of the MIMII dataset is utilized as the primary test object in this study. The purpose of this research is to find out which algorithm has a better ability to perform classification based on the loss value and accuracy after 300 epochs with a learning rate of 10^{-6}.

2. Materials and Methods

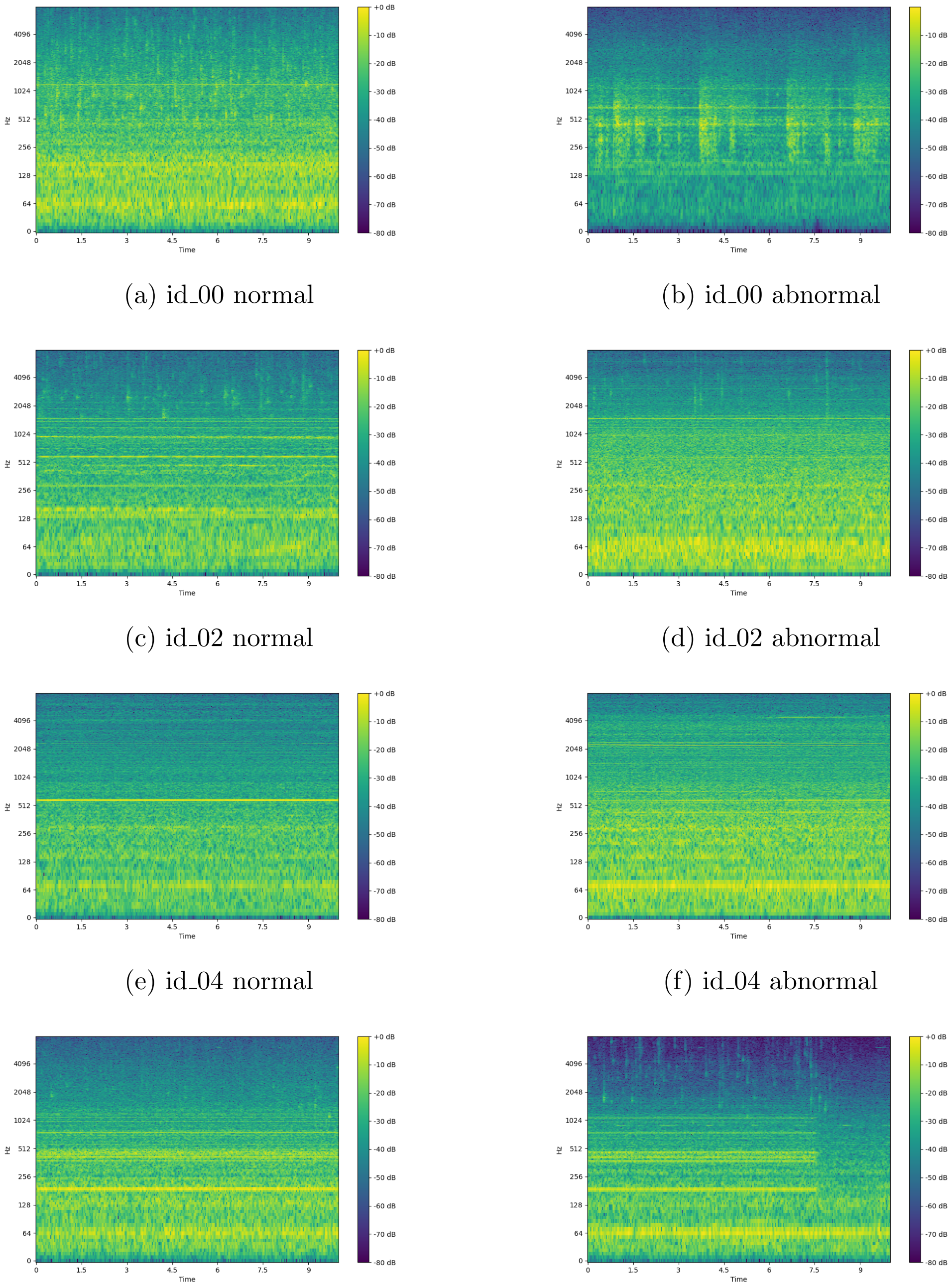
2.1. Materials

Table 1: Distribution of pumps audio files from the MIMII dataset

Device	**Model ID**	**Condition**	
		Normal	**Abnormal**
Pump	id_00	1006	143
	id_02	1005	111
	id_04	702	100
	id_06	1036	102

Purohit et al, a group of researchers at Hitachi Co. Ltd., released a dataset containing audio signal recordings under normal and abnormal operating conditions of four types of machines (fans, pumps, sliders, and valves) in 2019 [4]. The collection is called Malfunctioning Industrial Machinery Investigation and Inspection (MIMII) and is freely accessible. The audio data is in *.wav audio file format and includes machine sound mixed with real factory environment noise at -6 dB, 0 dB, and 6 dB signal-to-noise ratios. The abnormal condition audio data are not specifically labelled, yet they indicate various issues. The characteristics of the dataset are based on the type of equipment and signal-to-noise ratio(SNR), with each segment of the audio file being 10 seconds long and consisting of 160,000 samples at a sampling rate of 16,000 Hz and 16-bit resolution.

Journal of Physics: Conference Series **2673** (2023) 012032 doi:10.1088/1742-6596/2673/1/012032



(a) id_00 normal (b) id_00 abnormal

(c) id_02 normal (d) id_02 abnormal

(e) id_04 normal (f) id_04 abnormal

(g) id_06 normal (h) id_06 abnormal

Figure 1: Example spectrogram images of pump audio files under normal and anomalous conditions

In this study, a collection of pump sound recordings while operating under normal and abnormal conditions were used as test materials. These recordings are from the MIMII dataset and are freely accessible. Table 1 provides information on the allocation of pump sound files in the MIMII dataset with id_00, id_02, id_04, and id_06 representing the identifier(ID) codes of each pump. The table shows an unbalanced number of normal voice recordings and anomalous voice recordings. This can be a challenge for the classification algorithm in making a decision.

Journal of Physics: Conference Series **2673** (2023) 012032 doi:10.1088/1742-6596/2673/1/012032

Figure 1 shows examples of pump audio signal spectrogram shapes based on ID and operating conditions. The figure shows that normal and abnormal conditions can be distinguished based on the visualization of the acoustic signal. For example, in the comparison between normal and abnormal signals on the pump with ID id_00, the abnormal condition (Figure 1b) gives an acoustic signal that tends to be more dominant at high frequencies while the normal condition (Figure 1a) shows a dominant signal at low frequencies.

2.2. Methodology

In this research, the data used in each training and validation process is taken from a collection of pump audio signals contained in the MIMII dataset. The set is also grouped based on each machine ID. A total of 80% of the data is used as training data and the rest is for validation. The noise in the audio signal is assumed to be a form of augmentation of the original data, whereas, in the MIMII dataset, there are 3 types of noise, namely -6 dB, 0 dB, and 6 dB. Classification is based on the operation type of the acoustic signal under test without regard to ambient noise.

Figure 2 displays the configuration of the model simulation process using deep learning. As a first step before training, the mel-frequency cepstral coefficient of the acoustic signal will be extracted first and fed to the input part of the artificial neural network. In this study, the classification capabilities of four pre-trained deep-learning algorithms, namely AE, LSTM, GRU, and CNN, are compared using pump audio signal data from the MIMII dataset. Each algorithm will be trained using a learning rate of 10^{-6} for 300 epochs. The decision-making of the classification results is carried out in the last layer of the network using softmax activation which fulfils equation (1) for $i = 1, 2, \ldots, K$ and $\mathbf{x} = (x_1, x_2, \ldots, x_K) \in \mathbf{R}^K$. The performance of the algorithms is compared based on the accuracy and loss values of the validation results.

$$\sigma(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{1}$$

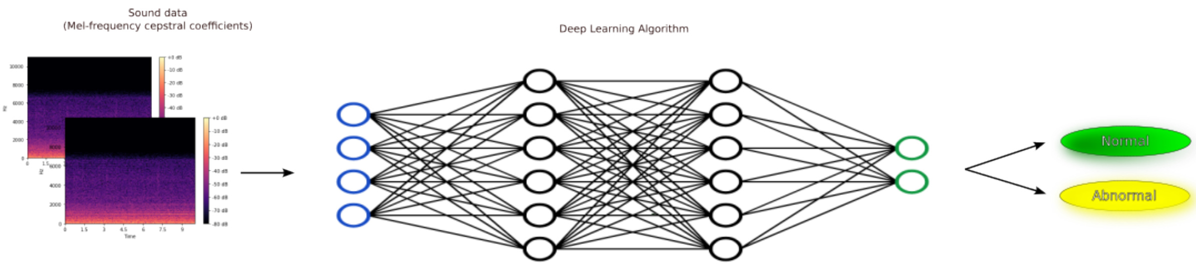


Figure 2: Illustration of neural network applied for classification

3. Results and Discussions

Figure 3 and 4 display the comparison of the accuracy and loss values of each algorithm based on the ID of each pump, respectively. The accuracy value produced by CNN is always above the accuracy values of other algorithms. The highest accuracy value is in the classification of acoustic signals from pump id_04 with 0.9917 and the lowest accuracy value is generated by the classification of acoustic signals from pump id_00 using LSTM and CNN. In terms of loss, CNN generates lower values for all pumps compared to AE, LSTM, and GRU. The lowest loss value is 0.3644 produced by the classification of acoustic signals from pump id_04 using CNN and the highest loss value is 0.6189 produced by the classification of pump id_04 with LSTM. The validation results show that each algorithm is capable of providing the ability to perform

Journal of Physics: Conference Series **2673** (2023) 012032 doi:10.1088/1742-6596/2673/1/012032

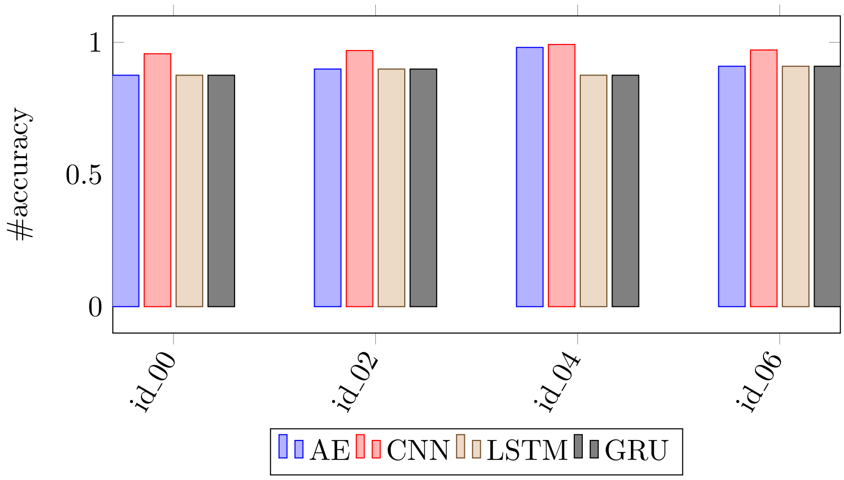


Figure 3: Comparison of accuracy values of each algorithm based on each machine ID

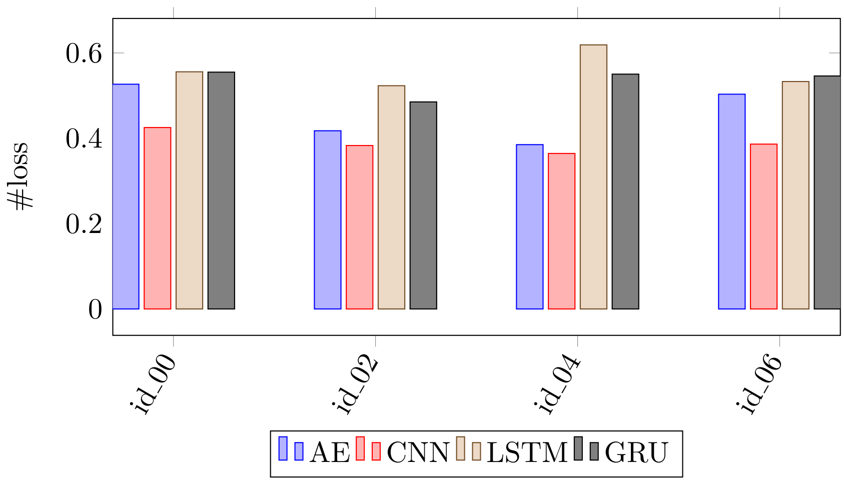
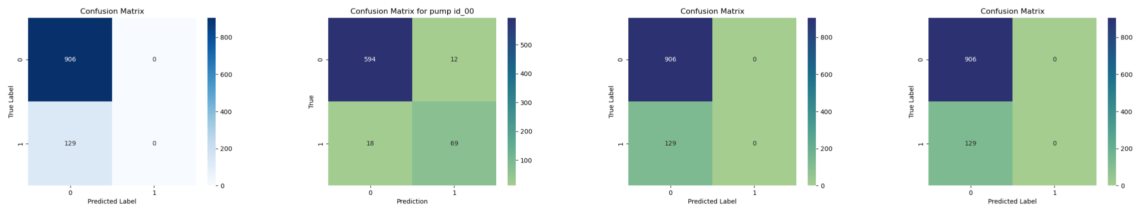
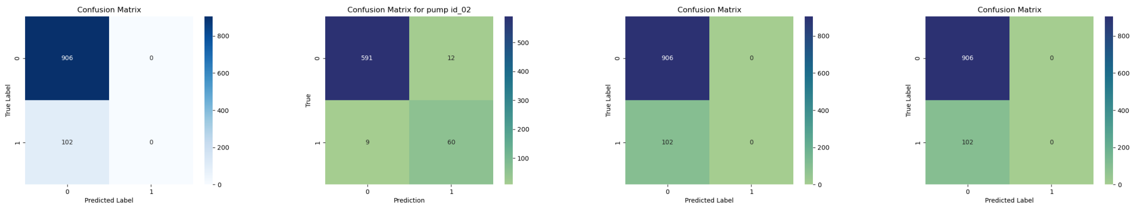


Figure 4: Comparison of loss values of each algorithm based on each machine ID



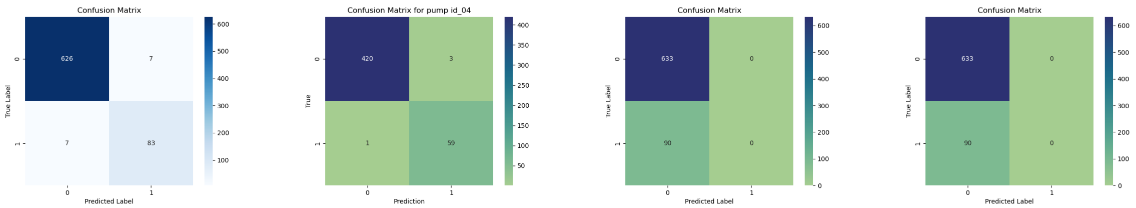
(a) id_00 with AE (b) id_00 with CNN (c) id_00 with LSTM (d) id_00 with GRU

Figure 5: Confusion matrices of pump id_00 classification



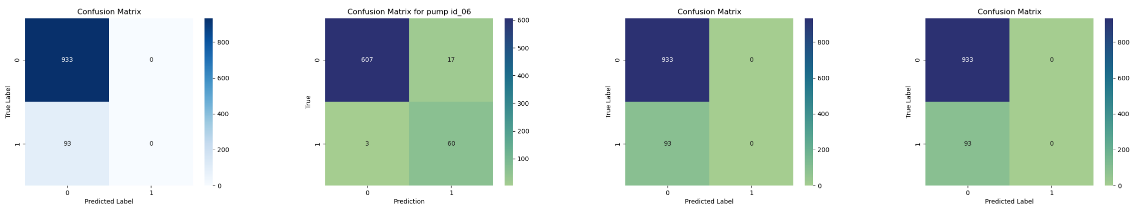
(a) id_02 with AE (b) id_02 with CNN (c) id_02 with LSTM (d) id_02 with GRU

Figure 6: Confusion matrices of pump id_02 classification



(a) id_04 with AE (b) id_04 with CNN (c) id_04 with LSTM (d) id_04 with GRU

Figure 7: Confusion matrices of pump id_04 classification



(a) id_06 with AE (b) id_06 with CNN (c) id_06 with LSTM (d) id_06 with GRU

Figure 8: Confusion matrices of pump id_06 classification

classification with good accuracy yet a sufficient amount of losses are still obtained.

Based on the test and validation results, a high accuracy value was obtained. However, at the same time, the resulting loss value is also quite high. This is due to the unbalanced amount of normal and abnormal condition data, according to the information contained in Table 1. Figure 5 to 8 are a collection of confusion matrices from the classification process based on the algorithm and pump ID. The figure shows that only the CNN algorithm is fully capable of differentiating normal and anomalous data based on the unbalanced data set. In addition, AE also produces the same ability in the case of the id_04 pump operation classification(Figure 7a). This is because the ratio of normal and anomalous data for the pump with id_04 is 7:1, which is lower if compared to other machines.

4. Conclusions

Based on the results of training and validation using the pump data contained in the MIMII dataset, the classification results with the highest accuracy and lowest loss are obtained when a convolutional neural network (CNN) is employed to generate the model. The highest accuracy value is obtained in the id_04 pump classification with a value of 0.9914. Meanwhile, the lowest loss is also generated in the same case with a value of 0.3644. In addition, CNN is also able to deal with classification problems with unbalanced data.

References

[1] Holford K, Eaton M, Hensman J, Pullin R, Evans S, Dervilis N and Worden K 2017 *Progress in Aerospace Sciences* **90**
[2] Cooper C, Wang P, Zhang J, Gao R, Roney T, Ragai I and Shaffer D 2020 *Procedia Manufacturing* **49** 105–111
[3] Tagawa Y, Maskeliunas R and Damaševičius R 2021 *Electronics* **10** 2329
[4] Purohit H, Tanabe R, Ichige K, Endo T, Nikaido Y, Suefusa K and Kawaguchi Y 2019 Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection URL `https://arxiv.org/abs/1909.09347`
[5] Jombo G and Zhang Y 2023 *Eng* **4** 47–79
[6] Shin J and Lee S 2023 *Electronics* **12** 409
[7] Ahn H and Yeo I 2021 *Sensors* **21** 5446
[8] Arronte Alvarez A and Gómez F 2021 *International Journal of Interactive Multimedia and Artificial Intelligence*
[9] Mobtahej P, Zhang X, Hamidi M and Zhang J 2022 *Computational and Mathematical Methods* **2022** 1–22
[10] Duan G, Su Y and Fu J 2023 *International Journal of Environmental Research and Public Health* **20** 1167
[11] Sams A and Zahra A 2023 *Bulletin of Electrical Engineering and Informatics* **12** 355–364
[12] Aqeel A, Hassan A, Khan M, Rehman S, Tariq U, Kadry S, Majumdar A and Thinnukool O 2022 *Sensors* **22** 1–15
[13] Kasthuri E and Balaji S 2021 *Materials Today: Proceedings*
[14] Fjellstrom C 2022 pp 3496–3504
[15] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning phrase representations using rnn encoder-decoder for statistical machine translation (*Preprint* `1406.1078`)
[16] Liu T, Jin Y, Wang S, Zheng Q and Yang G 2023 *Structural Health Monitoring* 147592172211435
[17] Tran T, Bader S and Lundgren J 2022 Denoising induction motor sounds using an autoencoder (*Preprint* `2208.04462`)
[18] Zhao Y, Hao H, Chen Y and Zhang Y 2023 *Electronics* **12** 2826
[19] Purohit H, Endo T, Yamamoto M and Kawaguchi Y 2022 Hierarchical conditional variational autoencoder based acoustic anomaly detection
[20] Wang Y, Chu Z, Ku I, Smith E C and Matson E T 2022 *2022 Sixth IEEE International Conference on Robotic Computing (IRC)* pp 186–189
[21] Zhen C and Changhui L 2021 *2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE)* pp 156–160
[22] Mahyub M, Souza L S, Batalo B and Fukui K 2022 *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)* pp 1–5
[23] Bold N, Zhang C and Akashi T 2019 *2019 International Conference on Cyberworlds (CW)* pp 85–88
[24] Valueva M, Nagornov N, Lyakhov P, Valuev G and Chervyakov N 2020 *Mathematics and Computers in Simulation* **177** 232–243 ISSN 0378-4754 URL `https://www.sciencedirect.com/science/article/pii/S0378475420301580`

Acknowledgment

The authors would like to express their sincere gratitude to the research team from Hitachi Ltd. for releasing the MIMII dataset and sharing it openly for research purposes: Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi.