A Study on Classification of Faulty Motor Sound Using Convolutional Neural Networks

Md Shafayet Jamil
Graduate School of Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Praveen Nuwantha Gunaratne
Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Hiroki Tamura
Faculty of Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan
Email: z322t01@student.miyazaki-u.ac.jp, ti20060@student.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

Classification of sound has its usage in various fields in today's world. In this paper we will go through the sound classification techniques for the detection of faulty machines with the help of the sound data produced by the machine. The focus is towards determining the pertinency of audio classification methods to detect faulty motors by their sounds; both in noisy and noise-free scenarios; so that the requirement of human inspection can be reduced in factories and industries. Noise reduction plays such an important role in improving accuracy of detection some researchers simulated data by adding noise for benchmarking their models. Hence noise reduction is widely used in audio classification tasks. Among various available methods, we have implemented an autoencoder for noise reduction. We have conducted the classification tasks on both noisy and denoised data using Convolutional Neural Network. Accuracy of classification on the data denoised using autoencoder is compared with the noisy ones. For classification, we used spectrogram, Mel-frequency cepstral co-efficient (MFCC) and Mel-spectrogram images. These processes yield promising results in distinguishing faulty motors by their sound.

Keywords: Faulty motor detection, autoencoder, convolutional neural network (CNN), audio classification

1. Introduction

Audio data is one of the most common multimedia types used in almost every sector of modern life throughout the past decades. The availability of audio recording devices has also been increased and the usage of audio data can be found in security surveillance systems, health monitoring systems, and various autonomous systems along with the inseparable usage in daily human life [1], [2], [3], [4]. Successful utilization of audio resources depends on the efficiency of classification (process of identification of pre-defined label for audio signals [5]), transcription, and perception of underlying contents through a bunch of transforming algorithms and machine learning approaches carried out on various sources of audio data, as in, voices, music, environmental sound, traffic sound etc. [1]. Our research aimed at distinguishing the faulty motors from the normal ones by means of the soundwave processing followed by training with convolutional neural network so that it can be used in factories to reduce the requirement for human inspection, and in preventing hazards. As the adequacy of classification often gets drastically affected by noise [6], we also have incorporated an autoencoder based noise reduction model in parallel to the classification process. CNN has many proven outcomes in the field of audio classification [7], [8], [9]. Convolutional neural networks have been used widely for classification since it has been found competent for both image classification e.g. vehicle classification [9] and sound classification [10] on various fields namely environmental sound classification (ESC) [11], animal sound classification [10], [12], vehicle classification [7], [13], emotion detection [14], violence detection [15] etc.

In this paper we will be showing results of applying Lenet-5 on image transforms of the audio signal. Lenet-5 was developed on the context of classification of images especially text [16]. However, while we can transform sound data into various image representations it can be an auspicious tool for classification of audios as well. We found approaches with applying CNN for the task of Environmental sound classification [11] and deploying Lenet for acoustic scene detection [17]. Despite the time consumption issues with large scale datasets CNNs are well suited for classification tasks within a few parameters [18]. For our research we have used Lenet-5 for classifying audios by means of transforming audios into feature images. We have used 3 types of feature-

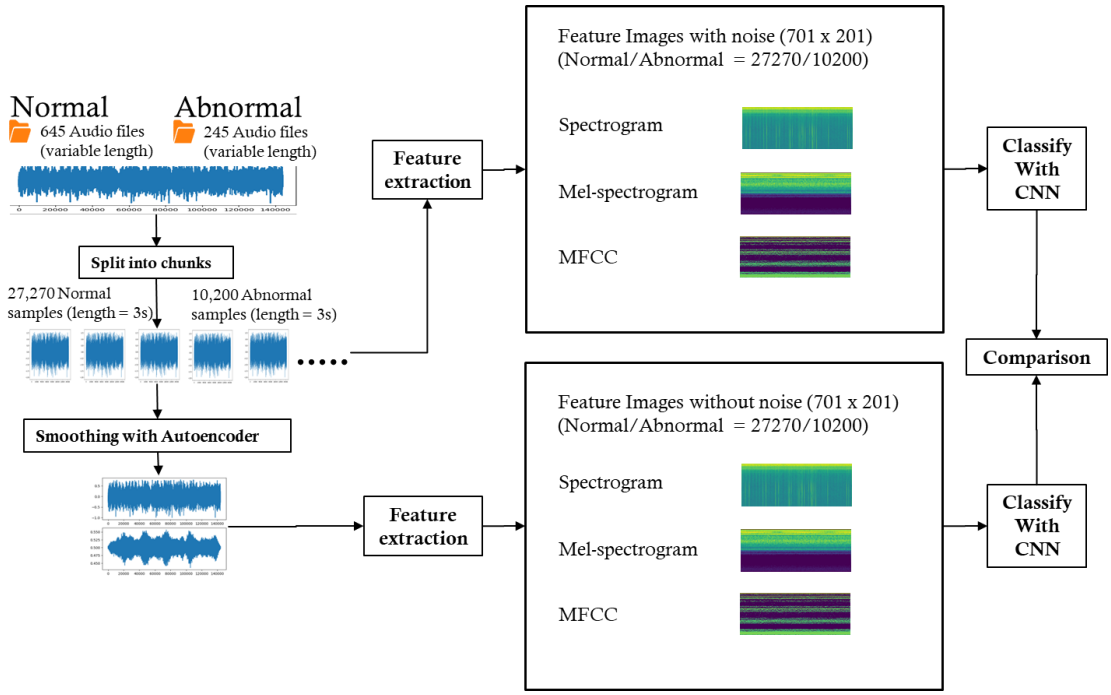


Fig. 1 Overview of faulty motor classification

images (Mel-Frequency Cepstral Coefficient, Mel-spectrogram, and Spectrogram) from both the noise-free original and smoothed (with an Autoencoder) audio chunks. Finally, we fed the classifier with both (noisy and smoothed) type of feature images separately and compared the resultant accuracy for noise association and the above mentioned 3 types of feature images as well. The process yields promising results. The whole process is shown in Fig. 1.

2. Dataset and Preprocessing

Our dataset is comprised of 890 samples of motor sounds, divided into two classes, 645 normal sound and 245 samples of abnormal sound; both are of variable length audio file in the form of '.wav' file at 48000Hz. The audio data was acquired using electronic stethoscopes. In preprocessing we split 3 seconds long chunks from the audio samples, having a length of 3 * 48000; and the samples which's lengths are less than 144000 (3 * 48000) were zero padded. Finally, each datapoint is normalized in such a way that the underlying values of the data stream contain real numbers ranging from -1.0 to 1.0 only.

3. Methodology

For classifying the uniform length audios of normal and abnormal motor sound; derived from the preprocessing; we divided the task into 2 standalone pipelines. The first one involves smoothing with autoencoder and the second one doesn't incorporate any noise reduction measure. In the later sections we have discussed these steps.

3.1. Noise reduction

Autoencoders are used for noise reduction in various audio recognition experiments [19]. An autoencoder is a composition of an encoder and a decoder block having convolutional and transpositional layers respectively. It can reconstruct the input without noise through a series of compressions and then decompressions [20]. We optimized the autoencoder model described in [20] by accompanying 4 convolutional layers and 4 transposition or deconvolutional layers as shown in Fig. 2. Also, in the transposition layers hyperbolic tangent activation (tanh) is used instead of PRelU for keeping the negative elements [21] in the resultant audios and to preserve a near-zero mean [22]. The results showing the smoothing with both types of activations are shown in Fig. 3 and Fig. 4 by superimposing the smoothed wave (yellow) onto the original one (blue).

Input
Encoder
Conv1D (128, k =7)
Conv1D (32, k =5)
Conv1D (16, k =3)
Conv1D (8, k=3)
Decoder
Transpose (8, k =3)
Transpose (16, k=3)
Transpose (32, k=5)
Transpose (128, k=7)
Output

Fig. 2 Autoencoder Model

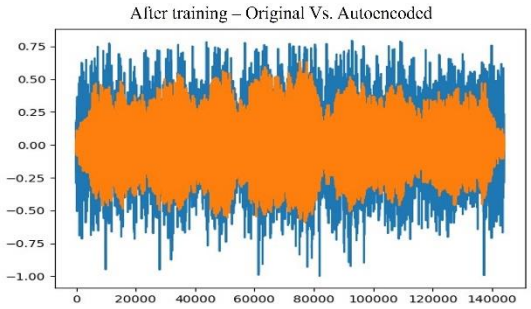


Fig. 3 Smoothing with autoencoder (tanh activation)

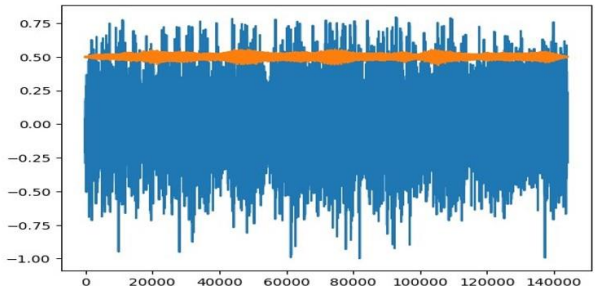


Fig. 4 Smoothing with autoencoder (PRelU activation)

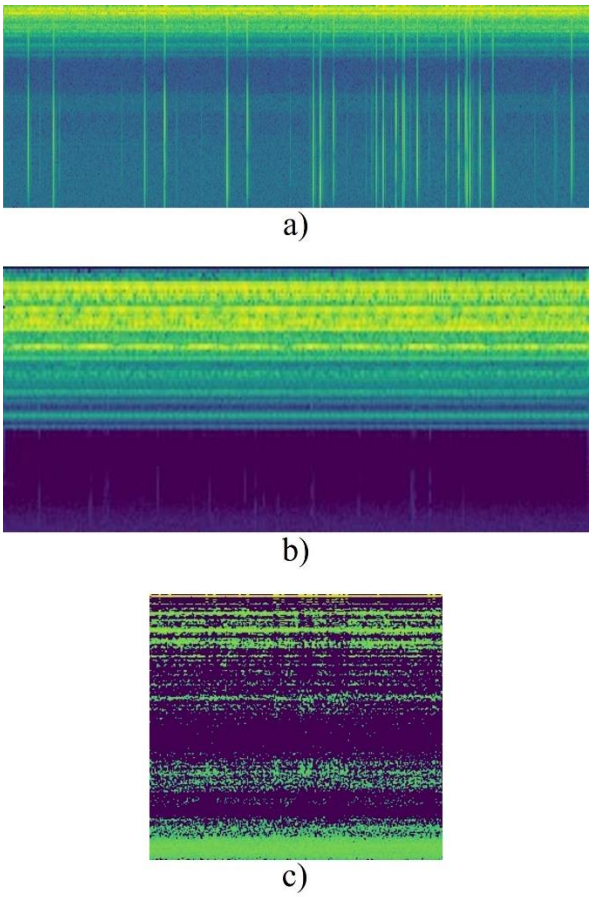


Fig. 5. Feature image samples a) Spectrogram b) Mel-spectrogram c) MFCC

3.2. Feature extraction

The selection of features has significant influence on classification accuracies [4]. That is why we fed both the noisy and smoothened audios through a transformation process which yield 3 types of feature images- a. Spectrogram, b. Mel-spectrogram, c. Mel-frequency cepstral coefficient (MFCC). We used a Python-based library called Librosa [23] for these processing.

a. Spectrogram

Transforming one dimensional audio signal into matrix like representation makes them suitable for training with neural networks [4]. Spectrograms are such visual depictions of frequency-wise strength of an audio which can improve classification results [24]. A sample spectrogram of a 3-second-long normal motor sound is shown in Fig 5 (a).

b. Mel-spectrogram

Mel scaled spectrograms can represent the signal analogous to the human auditory system. The formula for converting f Hz into m mel is $m = 2595 \log_{10}(1 + \frac{f}{700})$ [4]. The mel-spectrogram of a 3 second sound of abnormal motor looks like Fig. 5 (b).

c. Mel-frequency cepstral coefficient (MFCC)

MFCC incorporates Discrete Cosine Transform and has a compressed representation signal [4] so it is quite useful for training audio features. We have put an MFCC image of a smoothened normal motor sound (3 second) in Fig. 5 (c).

3.3. Classification

Image classification networks also have good results on sound data [24]. So, we trained these feature images with a CNN based model Lenet-5 (shown in Fig. 6) [16] which was proposed for classifying handwritten characters with requiring low preprocessing. Our implementation of Lenet-5 has an identical structure. It has 2 convolutional layers working as feature extractors followed by 3 fully connected layers working as classifier, as it appears in Fig. 7.

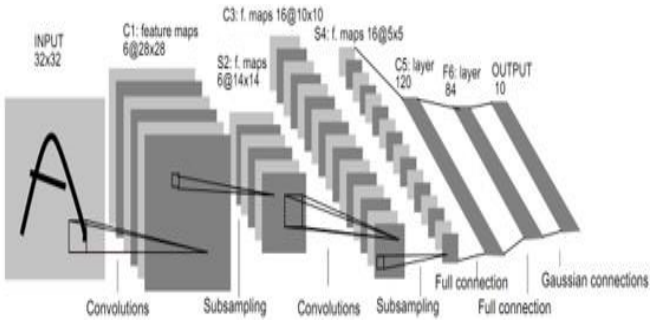


Fig. 6 Generic architecture of Lenet-5 [16]

Feature extractor
C1: Conv2d (3, 32, k=5)
S2: AvgPool2d (k=2)
C3: Conv2d (32, 64, k=5)
S4: AvgPool2d (k=2)
Classifier
C5: Convolution (128)
Gaussian Con.: Linear (84)
Output: Linear (2)

Fig. 7 Structure of the implemented Lenet-5 model.

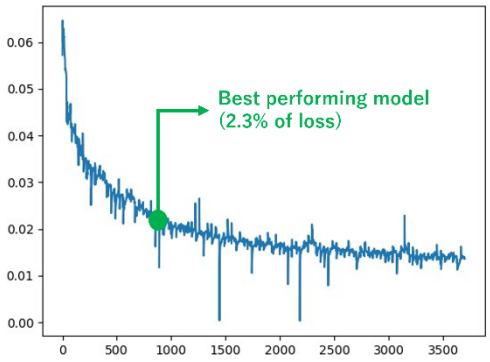


Fig. 8 Loss function of the autoencoder

4. Results

The results derived from the various experiments can be divided into 4 parts. Firstly, the performance of the autoencoder in terms of loss function (refers to Fig. 8). We have trained the autoencoder model up to 1.6% of loss using the metrics of mean squared errors (MSE), but the autoencoder model with 2.3% loss yield the best result during classification. Secondly, the straight-forward performance comparison between MFCC, Mel-spectrogram, spectrogram image-based classification of the model. Thirdly, comparison between the performance with and without using autoencoder smoothing, results of second and third steps are summarized in Table. 1. Fourthly, reviewing the impact of different dataset combinations, amounts data, and biases as shown in Table. 2. The best accuracies (%) found in these three steps are shown in Table 1 and Table 2. To put it in a nutshell, we can deduce that at least 97% accuracy could be expected from this approach in classifying 2 classes of motors (normal and abnormal) where a totally unknown type of motor sound is not present in the testing data. And as for noise reduction, it is not evident to have positive influence on classification accuracy with the current configuration of the model of the autoencoder that has been used for smoothing.

5. Conclusion

Sound classification with the help of CNN could be a feasible solution to various realistic problems especially in automation of maintenance and monitoring scenarios in large scale factories. It can reduce the need for manpower and improve Mean time to response (MTTR) by providing automated detection of faulty machines in industries. The scope of this paper could also be easily extended to other fields where a small number of classes are needed to be classified with high precision.

Table 1. Performance with and without autoencoder

Feature	With Autoencoder	Without Autoencoder
Spectrogram	98.39871898	98.71897518
Mel-spectrogram	99.93327996	99.90659194
MFCC	99.82652789	99.91638471

Table 2. Performance by dividing motors into two groups (motor 1-50 and motor 51-93)

Feature	Motor1–50 normal: Motor 51-93 abnormal	Motor51–93 normal: Motor 1-50 abnormal	Motor1–50 normal: Motor1-50 abnormal	Motor 51–93 normal: Motor 51-93 abnormal
Spectrogram	97.001	97.26	98.86	98.29
Mel-spectrogram	99.7	99.72	99.95	99.88
MFCC	99.9	99.9	99.95	99.99

References

1. L. Gao, K. Xu, H. Wang, and Y. Peng, "Multi-representation knowledge distillation for audio classification," Multimedia Tools and Applications, vol. 81, no. 4, pp. 5089–5112, Jan. 2022, doi: https://doi.org/10.1007/s11042-021-11610-8.
2. C. Clavel, T. Ehrette, and G. Richard, "Events Detection for an Audio-Based Surveillance System," IEEE Xplore, Jul. 01, 2005. https://ieeexplore.ieee.org/document/1521669
3. Y.-T. Peng, C. Lin, M. Sun, and K.-L. Tsai, "Healthcare audio event classification using Hidden Markov Models and Hierarchical Hidden Markov Models," Jun. 2009, doi: https://doi.org/10.1109/icme.2009.5202720.
4. M. Turab, T. Kumar, M. Bendechache, and T. Saber, "Investigating Multi-feature Selection and Ensembling for Audio Classification," International Journal of Artificial Intelligence & Applications, vol. 13, no. 3, pp. 69–84, May 2022, doi: https://doi.org/10.5121/ijaia.2022.13306.
5. Tuomas Virtanen, M. D. Plumbley, and D. Ellis, "Computational Analysis of Sound Scenes and Events." Springer, 2017, https://link.springer.com/book/10.1007/978-3-319-63450-0
6. J. Meyer, L. Dentel, and F. Meunier, "Speech Recognition in Natural Background Noise," PLoS ONE, vol. 8, no. 11, Nov. 2013, doi: https://doi.org/10.1371/journal.pone.0079279

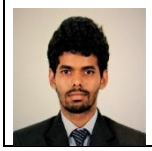
7. K. W. Cheng et al., “Spectrogram-based classification on vehicles with modified loud exhausts via convolutional neural networks,” Applied Acoustics, vol. 205, p. 109254, Mar. 2023, doi: https://doi.org/10.1016/j.apacoust.2023.109254.
8. C. Yang, X. Gan, A. Peng, and X. Yuan, “ResNet Based on Multi-Feature Attention Mechanism for Sound Classification in Noisy Environments,” Sustainability, vol. 15, no. 14, p. 10762, Jan. 2023, doi: https://doi.org/10.3390/su151410762.
9. M. A. Butt et al., “Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems,” Complexity, vol. 2021, pp. 1–11, Feb. 2021, doi: https://doi.org/10.1155/2021/6644861.
10. F. Merchan, A. Guerra, H. Poveda, H. M. Guzmán, and J. E. Sanchez-Galan, “Bioacoustic Classification of Antillean Manatee Vocalization Spectrograms Using Deep Convolutional Neural Networks,” Applied Sciences, vol. 10, no. 9, p. 3286, May 2020, doi: https://doi.org/10.3390/app10093286.
11. Z. Mushtaq and S.-F. Su, “Environmental sound classification using a regularized deep convolutional neural network with data augmentation,” Applied Acoustics, vol. 167, p. 107389, Oct. 2020, doi: https://doi.org/10.1016/j.apacoust.2020.107389.
12. Yin, Y., Tu, D., Shen, W., & Bao, J. , “Recognition of sick pig cough sounds based on convolutional neural network in field situations,” Information Processing in Agriculture, Nov. 2020, doi: https://doi.org/10.1016/j.inpa.2020.11.001.
13. M. A. Hedeya, A. H. Eid, and R. F. Abdel-Kader, “A Super-Learner Ensemble of Deep Networks for Vehicle-Type Classification,” IEEE Access, vol. 8, pp. 98266–98280, 2020, doi: https://doi.org/10.1109/access.2020.2997286.
14. R. Cai, L. Lu, H.-J. Zhang, and L. Cai, “Highlight sound effects detection in audio stream,” Jan. 2003, doi: https://doi.org/10.1109/icme.2003.1221242.
15. S. Pfeiffer, S. Fischer, and W. Effelsberg, “Automatic audio content analysis,” Proceedings of the fourth ACM international conference on Multimedia - MULTIMEDIA ’96, 1996, doi: https://doi.org/10.1145/244130.244139.
16. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998, doi: https://doi.org/10.1109/5.726791
17. Venkatesh Duppada and Sushant Hiray, “Ensemble Of Deep Neural Networks For Acoustic Scene Classification,” arXiv (Cornell University), Aug. 2017, doi: https://doi.org/10.48550/arxiv.1708.05826.
18. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, May 2012, doi: https://doi.org/10.1145/3065386.
19. S. Alharbi et al., “Automatic Speech Recognition: Systematic Literature Review,” IEEE Access, vol. 9, pp. 131858–131876, 2021, doi: https://doi.org/10.1109/access.2021.3112535.
20. K. Bajaj, D. K. Singh, and Mohd. A. Ansari, “Autoencoders Based Deep Learner for Image Denoising,” Procedia Computer Science, vol. 171, pp. 1535–1541, 2020, doi: https://doi.org/10.1016/j.procs.2020.04.164.
21. Swalpa Kumar Roy, S. Manna, Shiv Ram Dubey, and B. B. Chaudhuri, “LiSHT: Non-Parametric Linearly Scaled Hyperbolic Tangent Activation Function for Neural Networks,” arXiv (Cornell University), Dec. 2018, doi: https://doi.org/10.48550/arxiv.1901.05894.
22. S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “Activation functions in deep learning: A comprehensive survey and benchmark,” Neurocomputing, vol. 503, pp. 92–108, Sep. 2022, doi: https://doi.org/10.48550/arXiv.2109.14545.
23. B. McFee et al., “librosa: Audio and Music Signal Analysis in Python,” Proceedings of the 14th Python in Science Conference, 2015, doi: https://doi.org/10.25080/majora-7b98e3ed-003.
24. Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L., “Classifying environmental sounds using image recognition networks,” Procedia Computer Science, vol. 112, pp. 2048–2056, Jan. 2017, doi: https://doi.org/10.1016/j.procs.2017.08.250.

Authors Introduction

Mr. Md Shafayet Jamil

He received his bachelor’s degree in science in 2016 from the Institute of Information Technology, Jahangirnagar University, Bangladesh. He is currently a master’s student in University of Miyazaki, Japan.

Mr. Praveen Nuwantha Gunaratne



He received his Bachelor’s degree in Engineering in 2018 from the Faculty of Engineering, University of Moratuwa, Sri Lanka. He is currently a Doctoral student in University of Miyazaki, Japan

Prof. Hiroki Tamura

He received the B.E. and M.E. degree from Miyazaki University in 1998 and 2000, respectively. From 2000 to 2001, he was an Engineer in Asahi Kasei Corporation, Japan. In 2001, he joined Toyama University, Toyama, Japan, where he was a Technical Official in the Department of Intellectual Information Systems. In 2006, he joined Miyazaki University, Miyazaki, Japan, where he was an Assistant Professor in the Department of Electrical and Electronic Engineering. Since 2015, he is currently a Professor in the Department of Environmental Robotics. His main research interests are Neural Networks and Optimization Problems. In recent years, he has had interest in Biomedical Signal Processing using Soft Computing.