Acoustic Signal-Based Excavator Fault Detection Using Deep Learning Method

Yuhan Zhou[1], Foo Zhen Lin[1], Koh Yong Zhi[1], and Justin K.W. Yeoh[1]

[1]Department of Civil and Environmental Engineering, National University of Singapore, Singapore.

Abstract

As machinery assumes a critical role in modern construction, particularly in Singapore's development initiatives, maintaining excavators becomes paramount. Despite the prevalence of faults within this equipment, the scarcity of skilled mechanics compounds the challenge of timely diagnoses and maintenance. Leveraging deep learning methodologies, this research endeavors to analyze audio signals from excavators, aiming to identify distinctive patterns indicative of faults. Unlike existing studies primarily relying on vibration signals, this research focuses on audio signals for excavator fault prediction. Challenges involving ambient noise in construction sites and limitations in dataset size and imbalance compel the need for robust machine learning models capable of accurate fault diagnoses. The proposed methodology involves dataset collection, audio signal processing, feature extraction, and neural network training to differentiate normal operation from faulty conditions. This study delves into the application of machine learning and signal processing techniques to discern excavator conditions, aiming to classify their operational state as either faulty or operational. With an achieved 89.33% accuracy, 94.74% precision, and 85.71% recall, the method demonstrates promising performance. This research offers the potential to fortify excavator maintenance practices, potentially mitigating the impact of faults on construction productivity and costs.

Keywords –

Excavator Maintenance; Deep Learning; Fault Detection; Audio Signal Processing

1 Introduction

Throughout history, both equipment and labor have constituted essential components within the built environment [1]. The utilization of manual labor in construction has perennially remained a cornerstone. However, with technological advancements, the reliance on manual labor diminishes, giving way to a greater dependence on cutting-edge machinery at construction sites. These machines facilitate construction processes, fostering heightened productivity and long-term cost savings. Unlike humans, susceptible to fatigue, machinery operates tirelessly, accomplishing tasks equivalent to multiple human capacities without succumbing to exhaustion. Nevertheless, machines are susceptible to distinct forms of fatigue, manifesting as engine overheating or operational stress due to inadequate maintenance, leading to component faults.

This study aims to investigate excavators, extensively used across Singapore, particularly as the government endeavors to intensify subterranean development and coastal land reclamation. A prevalent challenge arises from the difficulty in instilling a sense of ownership and care for excavators. Primarily, excavator operators do not possess ownership, potentially impeding their understanding of how malfunctioning machinery impacts project outcomes, thus neglecting proper maintenance. Further compounded by demanding project schedules, these excavators endure heavy utilization, increasing susceptibility to faults necessitating servicing and upkeep. Unfortunately, only a select few excavator brands demonstrate prolonged durability with minimal maintenance costs.

Optimal excavator maintenance, beyond routine servicing, necessitates pre-empting potential faults. Minor issues such as hydraulic oil or radiator water leakage, or even loose bucket attachments, have the propensity to escalate into significant problems, potentially damaging the excavator if not promptly addressed. However, amidst a scarcity of foreign domestic workers and the industry's shift toward technologically advanced machines, the availability of experienced mechanics capable of accurately diagnosing and rectifying excavator faults diminishes. This makes manual diagnosis of excavator faults more difficult and troublesome.

To address this challenge, an effective approach involves deploying advanced machine learning techniques to autonomously discern machine activities or construction equipment by analyzing their distinctive sound patterns [1]. The progression of machine learning methodologies has facilitated fault detection successfully across diverse machinery, encompassing construction excavators. Typically, sensor fusion strategies, such as employing microphone arrays to capture machine-generated sounds, are utilized to gather audio signals. Even a single microphone holds the capability to capture sound from multiple machines.

While machine learning models leveraging sound signals have demonstrated success in domains like rotary machines [2] and combustion engines [3], their application specifically to excavators remains to be explored. Prior studies on excavator fault prediction [4], primarily relied on vibration signals as the primary input data. Contrary to vibration signals, audio signals are more intuitive and simpler to collect. However, the prevalent ambient noise within construction sites significantly undermines sound detection accuracy, complicating fault detection via sound signals. Therefore, establishing a dataset acquired in relatively quiet environments, encompassing both normal operating conditions and faults, becomes imperative as a standard resource for fault diagnosis.

Moreover, machine learning techniques rely on extensive, well-balanced datasets to construct highly accurate models. However, practical scenarios often present challenges in collecting such vast datasets. Instead, smaller and imbalanced datasets tend to be more common. Regrettably, limited research explores excavator fault detection efficacy using such data configurations. Hence, it becomes imperative to devise robust machine learning methodologies capable of effectively diagnosing excavator conditions, even when working with small and imbalanced datasets.

The primary objective of this research is to classify excavators as either faulty or operational through the utilization of signal processing methodologies in tandem with deep learning techniques. The process involves collecting faulty diagnosis datasets, denoising audio files, extracting audio features, and training a neural network to distinguish between normal operation and faulty conditions in excavators. Subsequently, the neural network's classification efficacy will be assessed using a validation dataset, employing pertinent evaluation metrics. An underlying assumption integral to this research involves acknowledging that the model's predictive capacity may not extend to anticipating excavator faults in their nascent or imminent stages, as the recordings of faults are already captured in their final state.

2 Related Work

2.1 Machine Fault Detection

The utilization of diverse sensor signals for machine fault detection has witnessed significant advancements in recent years. Leading researchers, such as Janssen & Arteaga (2020), have adeptly leveraged machine-learning methodologies to identify faults in rotary and structural machinery [5]. Signals play a crucial role in discerning activity and vibrations in large-scale mining and material handling equipment, where effective vibration management is paramount to mitigate the risk of potential mine collapses during ongoing excavation and movement. Wieckowski et al. (2020) utilized the Fast Fourier Transform (FFT) to convert vibration waveforms from the time domain to the frequency domain, subsequently devising a vibration control algorithm [6].

MEMS (Micro-Electro-Mechanical Systems)-based vibration sensors have gained widespread adoption across various applications, showcasing their effectiveness in accurately recognizing activity across a spectrum of equipment types [7]. However, the practical challenges associated with directly attaching these sensors to equipment, especially for smaller machinery or construction tools like concrete cutting saws and small concrete mixers, cannot be overlooked [8]. Additionally, the typical deployment of vibration sensors for individual machines presents limitations [9], particularly in scenarios where multiple machines operate concurrently. In contrast, an audio-based system offers a solution by obviating the need to attach a microphone to each machine, instead relying on strategically positioned microphones throughout the site.

Therefore, audio signals have emerged as valuable data resources for activity recognition and fault detection. Typically, this methodology adheres to a standard framework: (1) recording audio data using a single microphone, (2) utilizing FFT and Inverse Fast Fourier Transform (IFFT) to denoise the audio and extract pertinent features within specific time frames by transitioning between time and frequency domains, and (3) training a machine-learning algorithm using these extracted features [1]. Furthermore, pioneering techniques, such as employing a mobile microphone for spatial information in machine condition monitoring, have been explored [5]. The utilization of an array of multiple microphones has also proven effective in predicting the position of a mass on a vibrating plate, emulating structural flaws or engine imbalances. Its data processing involves employing peak-finding methods and three-dimensional imaging techniques. Collectively, these research findings underscore the efficacy of leveraging microphone signals as a valuable tool for fault detection and condition monitoring across a spectrum of diverse applications.

2.2 Sound-based Classification Using Neural Network

Acoustic sensors offer distinct advantages over other sensors due to their affordability and ease of placement, rendering them highly practical for event classification [10]. The collection of ambient sounds through simple microphones or sensors has paved the way for advanced signal-processing models. Consequently, there has been a surge in research focused on classifying construction work, machine types, and detecting faults using sound-based methodologies, driven by their cost-effectiveness and widespread applicability.

There have been studies in the literature focus on sound-based Construction Site Monitoring (CSM), aiming primarily at identifying activities or classifying brands and models of working machines. Multiple methodologies have been developed, often using Machine Learning (ML) approaches [11], [12]. Some commonly employed methods include Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs), and Gaussian Mixture Models (GMMs) [13].

Recently, deep learning methods using Neural Network become much more popular with their promising results. The most used model in the field of DL is the Convolutional Neural Network (CNN). Maccagno et al. employed a CNN-based model to classify different types and brands of machines on construction sites, obtaining 97.08% accuracy [14]. Similarly, Scarpiniti et al. presented a deep recurrent neural network (DRNN) model to classify five classes of multiple vehicles and tools using sound signals, reporting 97% accuracy [15]. While the mentioned studies exhibit promising advancements in classifying machine types and brands based on sound, an area that remains unexplored is the utilization of audio signals for fault detection. Despite the successes in identifying construction equipment and activities through sound analysis, there is a lack of research focused on leveraging these signals specifically for detecting faults or malfunctions within construction machinery such as excavators.

2.3 Imbalanced Data Processing

Conventional machine-learning techniques commonly rely on training models using balanced datasets, ensuring an equitable number of data samples for each class. Such datasets facilitate unbiased learning and streamline model performance evaluation. Many fault detection studies leverage extensive datasets encompassing several thousand samples for each class. However, datasets collected often exhibit imbalance, particularly when certain classes are rare, leading to limited samples representing these rare occurrences. Imbalanced datasets markedly affect model quality, resulting in poorer predictive performances for the minority class and potential overfitting to training samples, constraining the model's ability to generalize during testing [16]. This limitation is crucial, especially when the accuracy of predicting the minority class holds more significance than that of the majority class, particularly in scenarios where the cost of false negative predictions for the minority class is substantial [17].

Hence, various methods have emerged to address imbalanced datasets, primarily within the realm of data-level methods aiming to enhance the performance of standard machine learning algorithms. Resampling, a prevalent data-level technique, aims to rectify class distribution imbalances.

Random over-sampling is a widely used technique where, for instance, Hensman and Masko (2015) applied this approach to an image dataset by randomly duplicating minority class samples until reaching a balanced representation [18]. Results indicated that random over-sampling notably enhanced model performance, akin to that of a balanced dataset. An advanced technique, Synthetic Minority Over-sampling Technique (SMOTE), generates synthetic samples derived from the minority class data to alleviate overfitting issues typically associated with regular over-sampling [19]. While commonly used, SMOTE's efficacy has raised concerns. Sharma et al. (2018) indicated that in highly imbalanced classes (e.g., ratios around 1:1000), SMOTE-generated samples might negatively impact model performance by incorporating features overlapping with the majority class [20]. Elreedy and Atiya (2019) supported this observation, noting that due to SMOTE's nature, synthetic samples for very small or high-dimensional classes become less representative [21]. Nonetheless, despite its drawbacks, SMOTE improves classification, albeit not to the extent of a balanced dataset.

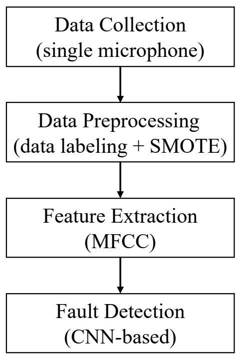


Figure 1. Overview of the proposed method

3 Method

Figure 1 presents an outline of the proposed method of diagnosing excavator conditions. Generally, the methodology comprises four key steps: a) data collection; b) data preprocessing; c) feature extraction; and d) fault detection. Detailed explanations of each of these steps follow in the subsequent sections.

3.1 Data Collection

Audio data from excavators was gathered using a single microphone to capture the noise emitted during their operation. Microphone placement differed based on whether the excavator was idle or in operation. For idle instances, the microphone was positioned in close proximity to the fault, while during operation, it was placed at a safe distance of 1 to 2 meters from the excavator. Given that most microphones possess 360° sensitivity [9], the utilization of a single microphone suffices for data collection in this study. However, for more complex on-site implementation scenarios, employing an array of consistently positioned microphones may represent a more effective strategy.

The study encompassed examinations of both “Good” and “Bad” conditioned excavators. “Good” noises represent well-conditioned excavator sounds without apparent faults, while “Bad” noises represent sounds indicative of faults. These faults encompass various issues such as hydraulic leakage in the arm, boom hose bursting, low engine power, control valve spool leaks, or visual indicators like smoke emanating from the excavator. Some faults manifest audibly, such as hydraulic leaks or hose bursts, while others are visually detectable, like smoke emissions.

3.2 Data Preprocessing

To enable the subsequent machine learning methodology, it was imperative to initially label the gathered dataset. Before audio recording, skilled mechanics conducted diagnostic assessments on the involved excavators. Their expertise and valuable insights into the excavators’ conditions were pivotal. Based on their evaluations, each file was meticulously labeled either as “0” denoting normal operational status or “1” indicating a faulty condition in the excavator. Furthermore, the reasons behind the faults were meticulously documented during this process.

Apart from data labeling, the imbalanced dataset is another issue. Neural network effectiveness is hindered by small or imbalanced datasets, which are common in machine learning due to challenges in acquiring large, representative samples. Addressing this, an enhanced oversampling technique, SMOTE, was employed on minority classes. SMOTE generates synthetic samples by interpolating minority class samples with their nearest neighbors of the same class, creating similar yet distinct synthetic samples. This mitigates overfitting risks, enhancing the model’s generalization capability [19].

3.3 Feature Extraction

Mel Frequency Cepstral Coefficient (MFCC) methodology was employed for audio signal feature extraction. Unlike the widely used Fast Fourier Transform, MFCC is designed to mirror human hearing capabilities. Human auditory perception of frequency increases logarithmically, meaning what’s perceived as a uniform frequency increase to humans isn’t linear. Additionally, human hearing is more adept at distinguishing lower frequencies compared to higher ones. MFCC efficiently captures unique features from recorded audio signals by employing the Mel scale, where lower-frequency coefficients possess greater spacing, while higher-frequency coefficients have narrower spacing, mimicking human hearing properties [22]. The Mel scale ensures that sounds equidistant on the scale are perceived as equally spaced by humans [23]. Leveraging MFCC, the model can learn crucial sound features indicative of faults.

3.4 Fault Detection

For fault detection in excavators, a binary classification approach - normal operation versus faulty - was undertaken. Utilizing MFCC spectrograms, specific spectrogram values at each point served as inputs. A proposed CNN-based neural network, depicted in Figure 2, was employed. Each convolutional layer utilized ReLU as the activation function. Max-pooling layers downsampled convolved features to save processing time and reduce image size. At each step, maximum values within each window were pooled into an output matrix. The model utilized binary cross-entropy as the loss function and the Adam optimizer. To optimize the performance of the proposed neural network, hyperparameter tuning was utilized.

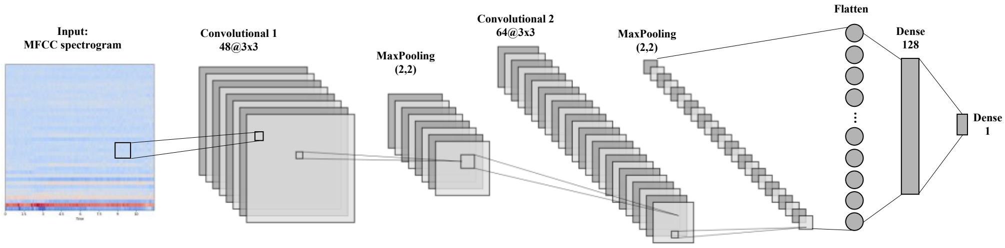
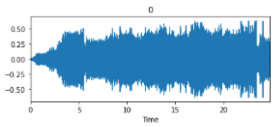


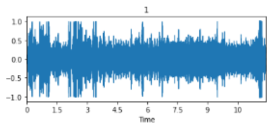
Figure 2. Proposed CNN structure

4 Validation

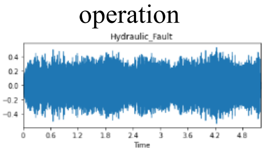
The dataset comprises 305 audio files of varied durations, spanning 1 to 88 seconds, recorded using a single microphone positioned according to the operational status of each excavator. Included in this dataset are recordings from diverse excavator brands such as Caterpillar (CAT), Sumitomo, Kobelco, Hitachi, Yanmar, Volvo, and Doosan. The excavators ranged in size from 5-tonne mini excavators to 38-tonne large excavators. Among the 305 files, 187 are classified as "good," indicating excavators in optimal working condition, while the remaining 118 files are categorized as "bad," representing excavators experiencing faults, which ranged from singular to multiple faults concurrently. These faults encompass various issues such as hydraulic leakages in excavator arms, burst boom hoses, low engine power, control valve spool leakages, smoke emissions, and others. While some faults, like hydraulic leakages, emit distinct sounds, others are only identifiable through visual cues like alarm lights or smoke emissions. Figure 3 shows some samples of the collected audio data, including excavators in normal operation and excavators with different types of faults.



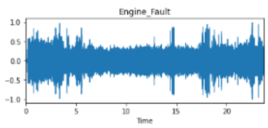
a) Excavator in normal operation



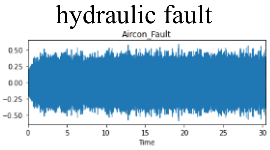
b) Excavator with fault



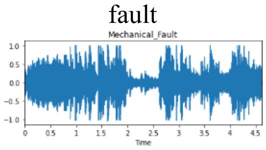
c) Excavator with hydraulic fault



d) Excavator with engine fault



e) excavator with aircon fault



f) Excavator with mechanical fault

Figure 3. Samples of the recorded excavator operations

After preprocessing, the original imbalanced dataset has been expanded to 374 samples, with 187 "good" noises and 187 "bad" noises. To validate the proposed method, the dataset is divided into two subsets: an 80% training dataset and a 20% test dataset. Evaluation of the proposed network's performance uses three key metrics: accuracy (see Equation (1)), precision (see Equation (2)), and recall (see Equation (3)). Accuracy denotes the ratio of correctly predicted observations (True Positives) to the total number of observations. Precision isolates the actual positive instances from the predicted positive dataset. Recall computes the count of true positives among all identified positives, including true positives and false negatives. Notably, in this study, recall holds more significance than the other metrics, as a false negative could significantly impact the excavator's lifespan if faults remain unidentified and unresolved.

$$accuracy = \frac{TP + TN}{Total\ Predictions} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

**TP: true positive; FP: false positive; TN: true negative; FN: false negative.*

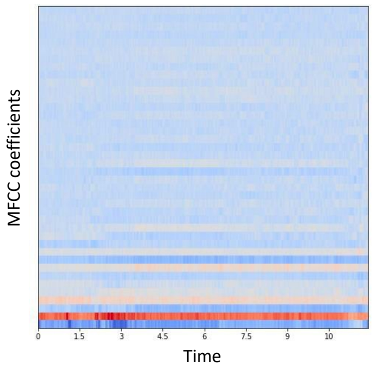


Figure 4. A sample of the MFCC spectrogram

Following the application of MFCC, the resulting MFCC spectrogram serves as the input for the proposed neural network. Illustrated in Figure 4 is a sample of the generated MFCC spectrogram. The X-axis delineates time, while the Y-axis represents the distinct MFCC coefficients. Additionally, it depicts frequencies transformed into Mel scale values, a logarithmic representation of signal frequencies. Each plotted point in the diagram signifies a specific MFCC at a precise moment in time. This spectrogram offers a visual portrayal of sound in the Mel scale, presenting an alternative to the frequency domain.

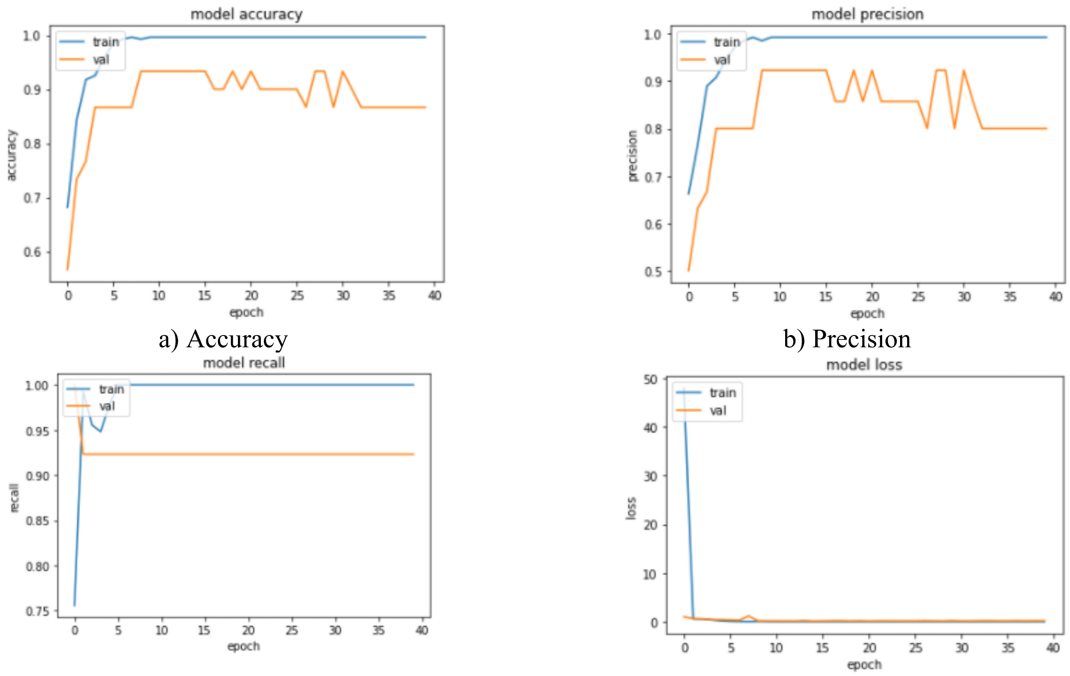
5 Results and Discussion

The findings presented in Table 1 demonstrate the model's good performance across both the training and test datasets. The training accuracy achieved 98.32%, while the test accuracy maintained a robust 89.33%. In terms of precision, the training and test sets displayed values of 97.35% and 85.71%, respectively, indicating the model's ability to minimize false positives, ensuring accurate positive predictions. Moreover, examining recall metrics reveals significant effectiveness. The training set exhibited a recall of 99.32%, while the test set maintained 94.74%. These results suggest the model's proficiency in capturing the majority of positive instances without missing many, highlighting its consistent performance on unseen test data.

The model's ability to generalize without overfitting is evident, showing consistent predictive power across various datasets. Figure 5 further illustrates the evolution of these metrics across epochs, depicting a rapid decrease in training and validation loss from the initial epochs, eventually converging to minimal values with a slight gap between the two losses. This presents an efficient learning ability of the proposed network.

The implementation of the proposed automatic fault detection method carries profound implications for machinery productivity and cost-effectiveness. By prioritizing the minimization of false-negative results, especially in the context of fault detection within excavators, this method significantly mitigates the risk associated with undetected faults. Such oversights can lead to operational disruptions, downtime, and potential safety hazards, all of which can incur substantial costs and impair machinery productivity.

By ensuring the timely identification and resolution of faults, the proposed method helps avert catastrophic failures that could result in extensive financial losses and pose risks to end-users. This proactive approach not only safeguards equipment integrity but also enhances operational efficiency by minimizing unplanned downtime and maintenance efforts. Moreover, the reliability, effectiveness, and accuracy demonstrated by this model underscore its capacity to optimize machinery performance and reduce operational expenses. By providing early and precise fault detection capabilities, this method enables maintenance interventions to be strategically planned, thereby minimizing costly repairs and maximizing equipment uptime.



a) Accuracy

b) Precision

c) Recall

d) Model loss

Figure 5. Results of the training and test process of the proposed CNN

In essence, the successful implementation of the proposed automatic fault detection method not only bolsters machinery reliability and safety but also translates into tangible benefits in terms of enhanced productivity and cost-effectiveness. Its robust performance positions it as a valuable asset for fault detection in excavators, offering significant potential for improving overall operational efficiency and profitability.

Table 1. Results of the proposed CNN

Dataset	Accuracy	Recall	Precision
Training	98.32%	99.32%	97.35%
Test	89.33%	94.74%	85.71%

6 Conclusion

This paper presents a CNN-based network designed for excavator fault detection using audio signals. The sound of excavator operations is captured through a single microphone, and expert mechanics assist in diagnosing the excavators' statuses. Both normally functioning excavators and those with faults are recorded, with subsequent data labeling based on the mechanics' assessments. To address dataset imbalance, the SMOTE oversampling method is employed to balance the dataset. The extraction of frequency features from audio signals is accomplished through the generation of MFCC spectrograms. These spectrograms serve as input for the proposed CNN-based network. The evaluation of the network's performance employs accuracy, precision, and recall metrics, demonstrating good generalization and robustness in excavator condition recognition.

This study contributes to leveraging machine learning applications to improve excavator serviceability. The COVID-19 pandemic highlighted Singapore's heavy reliance on foreign manual labor, resulting in significant downtime and project slowdowns. There is a growing consensus that a paradigm shift in site operations is necessary, especially considering the migration of talent to other sectors, potentially reducing expertise in this industry. Implementing machine learning could alleviate dependency on senior mechanics for fault identification. It could empower on-site engineers or workers to recognize fault-related sounds promptly, enabling immediate servicing to prevent excavator breakdowns.

A limitation of this study lies in the range of its dataset. As previously mentioned, audio data capturing excavator operations was acquired using a single microphone within a relatively quiet environment with minimal ambient noise. There is merit in extending the dataset to encompass a diverse range of real-world conditions, particularly those characterized by higher levels of background noise, to further validate the proposed method. Apart from gathering data from real-world conditions, acquiring a comprehensive range of normal operation sounds holds significant value in enriching the dataset. This augmentation can enhance the rationality of fault diagnosis and provide additional support for anomaly detection, which presents itself as a potential avenue for future research exploration. Additionally, the data labeling process relied on expert input, which can be time-consuming and labor-intensive, particularly as the dataset expands. This process could be streamlined by implementing automated or semi-automated labeling procedures.

In terms of future research directions, a more granular classification of faults into major and minor categories could yield substantial insights. Moreover, an in-depth analysis of the network's discriminatory capabilities across varied fault scenarios is warranted. Additionally, expanding efforts to correlate specific faults with their respective excavator brands, leveraging an augmented dataset, could enhance the network's proficiency in accurately identifying excavator-specific faults. The inclusion of datasets spanning a broader spectrum of excavator models holds promise for enriching research outcomes.

References

[1] B. Sherafat *et al.*, "Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review," *Journal of Construction Engineering and Management*, vol. 146, no. 6, p. 03120002, Jun. 2020, doi: 10.1061/(ASCE)CO.1943-7862.0001843.

[2] O. Janssens *et al.*, "Convolutional Neural Network Based Fault Detection for Rotating Machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, Sep. 2016, doi: 10.1016/j.jsv.2016.05.027.

[3] L. Barelli, G. Bidini, C. Buratti, and R. Mariani, "Diagnosis of internal combustion engine through vibration and acoustic pressure non-intrusive measurements," *Applied Thermal Engineering*, vol. 29, no. 8, pp. 1707–1713, Jun. 2009, doi: 10.1016/j.applthermaleng.2008.07.025.

[4] Q. Zhou, G. Chen, W. Jiang, K. Li, and K. Li, "Automatically Detecting Excavator Anomalies Based on Machine Learning," *Symmetry*, vol. 11, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/sym11080957.

[5] L. A. L. Janssen and I. Lopez Arteaga, "Data processing and augmentation of acoustic array signals for fault detection with machine learning," *Journal of Sound and Vibration*, vol. 483, p. 115483, Sep. 2020, doi: 10.1016/j.jsv.2020.115483.

[6] J. Więckowski, W. Rafajłowicz, P. Moczko, and E. Rafajłowicz, "Data from vibration measurement in a bucket wheel excavator operator's cabin with the aim of vibrations damping," *Data in Brief*, vol. 35,

p. 106836, Apr. 2021, doi: 10.1016/j.dib.2021.106836.

[7] B. Sherafat, A. Rashidi, Y.-C. Lee, and C. R. Ahn, "A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment," *Sensors*, vol. 19, no. 19, Art. no. 19, Jan. 2019, doi: 10.3390/s19194286.

[8] B. Sherafat, A. Rashidi, and S. Asgari, "Sound-based multiple-equipment activity recognition using convolutional neural networks," *Automation in Construction*, vol. 135, p. 104104, Mar. 2022, doi: 10.1016/j.autcon.2021.104104.

[9] C.-F. Cheng, A. Rashidi, M. A. Davenport, and D. V. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," *Automation in Construction*, vol. 81, pp. 240–253, Sep. 2017, doi: 10.1016/j.autcon.2017.06.005.

[10] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, "Acoustic Scene Classification: Classifying environments from the sounds they produce," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 16–34, May 2015, doi: 10.1109/MSP.2014.2326181.

[11] S. Ahmad *et al.*, "Environmental sound classification using optimum allocation sampling based empirical mode decomposition," *Physica A: Statistical Mechanics and its Applications*, vol. 537, p. 122613, Jan. 2020, doi: 10.1016/j.physa.2019.122613.

[12] S. Scardapane, M. Scarpiniti, M. Bucciarelli, F. Colone, M. V. Mansueto, and R. Parisi, "Microphone array based classification for security monitoring in unstructured environments," *AEU - International Journal of Electronics and Communications*, vol. 69, no. 11, pp. 1715–1723, Nov. 2015, doi: 10.1016/j.aeue.2015.08.007.

[13] R. V. Sharan and T. J. Moir, "An overview of applications and advancements in automatic sound recognition," *Neurocomputing*, vol. 200, pp. 22–34, Aug. 2016, doi: 10.1016/j.neucom.2016.03.020.

[14] A. Maccagno, A. Mastropietro, U. Mazziotta, M. Scarpiniti, Y.-C. Lee, and A. Uncini, "A CNN Approach for Audio Classification in Construction Sites," in *Progresses in Artificial Intelligence and Neural Systems*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds., in Smart Innovation, Systems and Technologies. , Singapore: Springer, 2021, pp. 371–381. doi: 10.1007/978-981-15-5093-5_33.

[15] M. Scarpiniti, D. Comminiello, A. Uncini, and Y.-C. Lee, "Deep Recurrent Neural Networks for Audio Classification in Construction Sites," in *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, pp. 810–814. doi: 10.23919/Eusipco47968.2020.9287802.

[16] Z. Li, K. Kamnitsas, and B. Glocker, "Analyzing Overfitting Under Class Imbalance in Neural Networks for Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1065–1077, Mar. 2021, doi: 10.1109/TMI.2020.3046692.

[17] S. Vucetic and Z. Obradovic, "Classification on Data with Biased Class Distribution," in *Machine Learning: ECML 2001*, L. De Raedt and P. Flach, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2001, pp. 527–538. doi: 10.1007/3-540-44795-4_45.

[18] D. Masko and P. Hensman, "The Impact of Imbalanced Training Data for Convolutional Neural Networks," 2015. Accessed: Dec. 15, 2023. [Online]. Available: https://www.semanticscholar.org/paper/The-Impact-of-Imbalanced-Training-Data-for-Neural-Masko-Hensman/62e81797fff75603a3d7c7759e6efac4fd2b6b31

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *jair*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[20] S. Sharma, C. Bellinger, B. Krawczyk, O. Zaiane, and N. Japkowicz, "Synthetic Oversampling with the Majority Class: A New Perspective on Handling Extreme Imbalance," in *2018 IEEE International Conference on Data Mining (ICDM)*, Nov. 2018, pp. 447–456. doi: 10.1109/ICDM.2018.00060.

[21] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Information Sciences*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.

[22] A. Meghanani, A. C. S., and A. G. Ramakrishnan, "An Exploration of Log-Mel Spectrogram and MFCC Features for Alzheimer's Dementia Recognition from Spontaneous Speech," in *2021 IEEE Spoken Language Technology Workshop (SLT)*, Shenzhen, China: IEEE, Jan. 2021, pp. 670–677. doi: 10.1109/SLT48900.2021.9383491.

[23] A. Mahmood and U. Köse, "Speech recognition based on convolutional neural networks and MFCC algorithm," *Adv. Artif. Intell. Res.*, vol. 1, no. 1, Art. no. 1, Jan. 2021.