The European Journal of Research and Development

journals.orclever.com/ejrnd

OR CLEVER Science & Research Group

Review Article

A CNN Based Ensemble Approach for Malfunction Detection from Machine Sounds

Berk Öztürk[1*], Melih Yılmaz Öğütcen[2*], Berk Kayı[3*], Mehmet Ali Varol [4*], Tayfun Özçay[5*], Nermin Yalçı[6*], Esra Akca[7*], Yasin Dinç[8*], Semra Erpolat Taşabat[9*]

[1] Linktera, Turkey, ORCID: 0000-0002-8548-5574, berk.ozturk@linktera.com
[2] Linktera, Turkey, ORCID: 0000-0002-6143-3112, melih.ogutcen@linktera.com
[3] Linktera, Turkey, ORCID: 0000-0003-1623-8120, berk.kayi@linktera.com
[4] Linktera, Turkey, ORCID: 0000-0001-8028-6307, mehmetali.varol@linktera.com
[5] Borusan CAT, Turkey, ORCID: 0000-0003-0011-7877, tozcay@borusan.com
[6] Borusan CAT, Turkey, ORCID: 0000-0001-6444-4602, nyalci@borusan.com
[7] Borusan CAT, Turkey,ORCID: 0000-0003-3790-8584, esakca@borusan.com
[8] Borusan CAT, Turkey,ORCID: 0000-0002-8692-1091, ydinc@borusan.com
[9] Mimar Sinan Fine Arts University, Turkey, ORCID: 0000-0001-6845-8278, semra.erpolat@msgsu.edu.tr

(First received March 22, 2022 and in final form June 04, 2022)

Reference: Akca, E., Özçay, T., Dinç, Y., Yalçı, N., Erpolat Taşabat, S., Varol, M. A., Kayı, B., Öğütcen, M. Y., & Öztürk, B. A CNN Based Ensemble Approach for Malfunction Detection from Machine Sounds. *The European Journal of Research and Development*, *2*(2), 411–420

Abstract

Together with the meaning and essence of data for the company nowadays; The variety of data also differs. One of these differentiating data types is sound. Borusan Makina ve Güç Sistemleri A.Ş. the data obtained from the Caterpillar construction machines of. The machine sound gives clues about many malfunctions. Artificial intelligence systems of the heard sound will be integrated into business processes. Every tone can be converted. With this, the properties and estimates of the sound grids are used. In this direction; While the incident is getting in the way of his business, an unfortunate project occurs with a similar visitor. The traditional will use a meaningful method by listening to the producer's sound and technology and innovation to develop easy blueprints of decisions that cannot be diverted to sound data. Thanks to the real-time model with short-term audio recording, it is instantly predicted whether there is a problem in the machine. Free from personal and technical comments; By examining the patterns of sound waves, it is aimed to be made without cancellation.

Keywords: Sound Diagnostic, Sound Classification, Convolutional Neural Network, Ensemble Model

1. Introduction

Knowing when a machine will fail in a business is very important in terms of customer satisfaction, the competitiveness of the business, and taking measures to prevent losses before they occur. To prevent these damages and to get the maximum benefit from their machines, businesses try to prevent untimely machine failures of their machines by performing daily, monthly, annual, and periodical maintenance-repair and part replacements. However, unfortunately, due to some unforeseen reasons in the working process, machines can sometimes cause problems. Considering these situations, the company was asked to continue its activities without interrupting its operations by utilizing the results of the failure estimation method. Repair maintenance time and resources must be used efficiently to eliminate faults in a short time, reduce costs, and ensure service quality and continuity. For this purpose, it is aimed to provide convenience to the customer and increase customer satisfaction by estimating the malfunction from the sound data. In the literature research, fault detection with sound was not found in the construction equipment sector.

The characteristics of the sound signal generated by a mechanical system can be represented by a typical sound waveform. The typical waveform is associated with the state of operation for the monitored components within the device. Thus, signal-processing techniques can provide useful methods for fault diagnosis and condition monitoring. Extracting information or features that are closely related to a specific fault is a great challenge in fault diagnostic and condition monitoring based on sound signal-processing techniques [1].

In this study, fault recognition from sound through artificial intelligence consists of 2 stages. Stage 1 is on the prediction model's understanding that it is the sound of construction equipment. At this stage, the KNN algorithm was used. After the estimation results that passed the 1st Stage, the 2nd Stage is; It is about determining whether the sound recording of the work machine is defective. At this stage, an ensemble model structure was developed by defining the outputs of the Convolutional Neural Network (CNN) model as input to the XGBoost algorithm. This established model structure was integrated into the company's mobile application, end-to-end; By taking a short-term sound recording, and passing through the model stage in the background, the end-user is notified with a message expressing the current status of the work machine.

2. Methods

2.1.Preprocessing

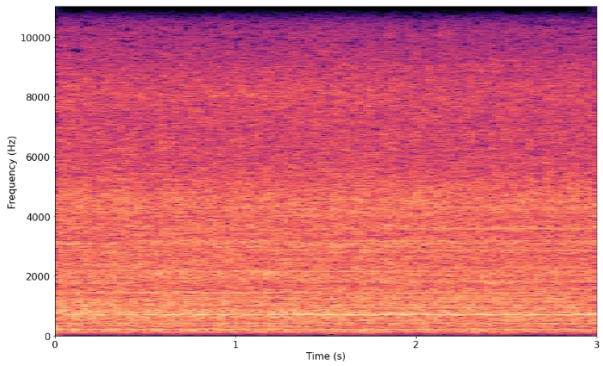


Fig 1. A spectrogram image of a 3 second sound (frequency is represented linearly on the y-axis)

The spectrogram in Figure 1 is of a 3-second machine sound. In this spectrogram image, the signal is distributed over a wide frequency range. A method similar to Welch's method was used to examine the frequency-energy distribution in the entire data set. In this method, the signal is divided sequentially with a certain overlap ratio, and the FFTs (Fast Fourier Transform) of each obtained segment are taken. By summing these FFTs, an overall frequency-energy profile of the signals is obtained. Since the sounds are divided into sections, the effect of the noise is reduced on the overall profile. In this method, the sounds are divided into 3 seconds and the overlap amount of 1 second (~33.3% overlap ratio) is determined. The resulting profile is shown in Figure 2.

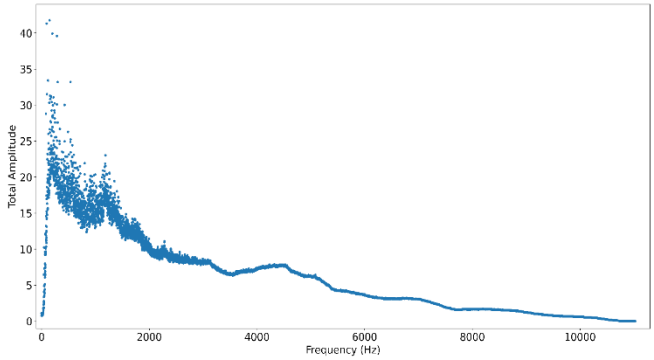


Fig 2. General frequency-amplitude profile of all sounds obtained with Welch's Method

When the profile in Figure 2 is examined, the energy distribution of sounds below 50Hz and above 9kHz decreases. This shows that the general profile of all sounds is also distributed over a wide frequency range.

2.2.Sound Outlier Detection

The sounds in the dataset are collected by the technicians in the field with a mobile application developed during the maintenance of the machine. These collected sounds may include noises such as wind and speech sounds. Correct detection of sounds as outliers is important for the accuracy of the classification model. Therefore, various sounds such as white noise, traffic, speech, and music sounds were collected from YouTube. Since these sounds do not belong to machine sounds, they are labeled as outliers. 120 clean machine sounds listened to manually are also labeled as machine sounds. First, these sounds were divided into 5-second segments non-overlap. The MFCC coefficients of 20 lengths of each separated segment are averaged in the time domain and a KNN model is trained with these features. Due to the success of the KNN model on low-dimension data, a sharp separation was achieved for outlier detection. If more than 50% of the collected sound is an outlier, it is not included in the model for a healthy evaluation.

2.3.Noisy Segment Detection

The outlier detection model mentioned in section 2.2 removes the sound from the model if more than 50% of the sound is an outlier. Some of the sounds may contain noisy segments and an approach similar to the model in section 2.2 is used to separate these segments. In this section, instead of segments of 5 seconds, the sounds are divided into 3 seconds and a 1-second overlap amount (~33.3% overlap ratio) is used. Likewise, MFCC coefficients were obtained and the KNN model was trained. In this way, noisy segments in the collected sound do not enter the model at all.

3. Experiments

3.1. Datasets

Sounds of construction machines are recorded via the company mobile application; It is collected by technicians and customers and stored in the cloud infrastructure. Through this application, sounds were collected from work machines for a minimum of 15 seconds and a maximum of 1 minute by the technicians and customers in the service. After the diagnostics, the malfunction of the machine was determined and the type of malfunction was combined with the recorded sound of the application. Fault types are divided into 4

different classes Fuel Systems, valves, Handle Beds and Others. The distribution of the data set is as follows.

Table 1. Type of malfunctions and distribution of the dataset

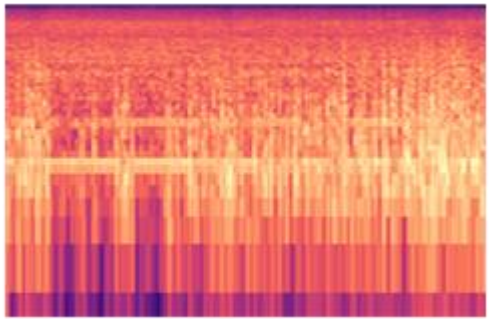
Type of Malfunction	Count	Merged Classes	Total Count	Number of segments
Healthy	169	Healthy	169	4463
Fuel System	42	Unhealthy	65	1437
Other	11			
Valves	9			
Handle Beds	3			

The data set described in Table 1 shows an unbalanced distribution, especially in Valves and Handle Beds classes, the lack of data is an important reason why we consider the problem as a binary classification problem that will enable us to achieve high performance. To make the feature to be extracted from the sounds more comprehensive and to ensure that the sounds can be compared under equal conditions, the sounds were divided into segments of 3 seconds and the overlap amount of 1 second was determined as ~33.3% overlap ratio. After this process, HPSS features for deep learning architectures and structural data for the XGBoost algorithm, which will be mentioned in section 3.3, are used.

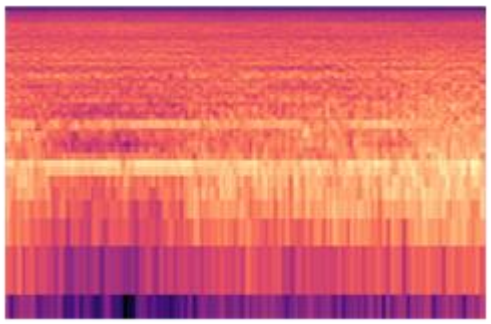
3.2. Convolutional Neural Network (CNN) Based Model

Harmonic and percussive components obtained by the source separation method from the Mel spectrogram and Mel spectrogram are used as features (HPSS) [2]. In this process, energy is assigned to each time-frequency bin according to the higher response of the horizontal (harmonic) and vertical (percussion) filter at that location. In this way, periodic sounds can be represented more clearly in the Harmonic component, and percussive sounds with high bandwidth can be represented more clearly. As seen in Figure 3, for a 3-channel input, the first channel is Mel spectrogram, the second channel is Harmonic components, and finally, the third channel is Percussive components.

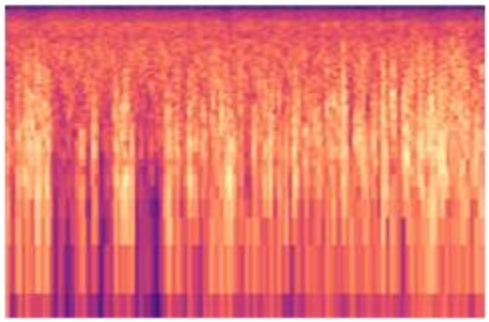
EfficientNet is a lightweight convolutional neural network architecture with high performance and few parameters on different datasets such as ImageNET used for transfer learning [3]. In this study, HPSS features are extracted for 3-second segments with 1-second overlap, and each feature is trained with 3 different CNN models: the CNN architecture shown in Figure 4, EfficientNetB0, and EfficientNetB7. Probability values from this model are used as an attribute in the XGBoost model.



(a)



(b)



(c)

Fig 3. Mel Spectrogram (a), Harmonic components (b), and Percussive Components (c) of a 3-second sound (frequency is represented logarithmically on the y-axis).

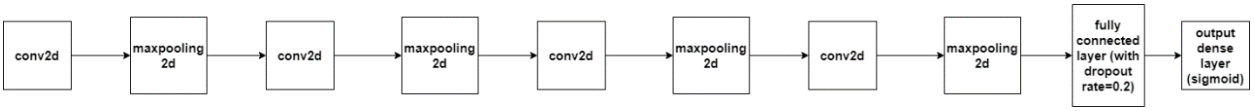


Fig. 4. CNN architecture

In image classification problems, in cases where classes are unbalanced, data augmentation methods such as data augmentation are used to achieve balance. In these methods, the image can be zoomed, rotated, cropped, etc. Various combinations of methods are created and used in the model. However, in the spectrogram images, the x-axis represents the time and the y-axis represents the frequency. Any action to be taken may distort the characteristic structure of the image on which the sound is represented and lead to misinterpretation. Due to the unbalanced classes in this model, the data belonging to the Unhealthy class were oversampled and trained in the models without any additional processing in the train set.

3.3. XGBoost Based Model

In the study in [4], researchers trained a BP-NN (backpropagation neural network) with three types of structured features for CNC machines, namely amplitude, frequency, and power (wavelet). With a similar approach to this approach, in this study, 4 types of structured features in Table 2, namely amplitude, frequency, power, and MFCC, were created. Different from the study in [4], Mel frequency cepstral coefficients of n=20 length, probability values obtained from the CNN model, and counter-information including the cumulative working time of the machine were added as an attribute to these features.

Table 2. Structured Feature List

Type	Sound Signal Feature
Amplitude	Max amplitude
	Mean amplitude
Frequency	Max frequency peak
	Second harmonic peak
	Third harmonic peak
	Sum of max and second peak
	Sum of max, second and third peak
Power	A5 frequency coefficient
	D5 frequency coefficient
	D4 frequency coefficient
	D3 frequency coefficient

	D2 frequency coefficient D1 frequency coefficient
MFCC	Mean MFCC coefficients (n=20)
Probability	Prediction probability of CNN
Type	Machine Feature
Counter	Total working hours of machine

The probability outputs produced by the CNN model are used as a feature in the structured data. This allows us to evaluate the state of the machine with different attributes besides the sound. A machine with high total uptime is also more likely to fail, and using this information as an instant feature increases the performance of our model. XGBoost is a widely used efficient and high-performance gradient boosting algorithm [5]. With the features in Table 2, the XGBoost model is e1-second and a final model of the general structure we have created has been obtained. The flow diagram of our ensemble approach, which we propose in this study, is shown in Figure 5.

4. Results

The data set for the ensemble structures we created is divided into 5 different distributions. HPSS attributes for each fold are trained in CNN architecture, EfficientNetB0, and EfficientNetB7 architectures. Obtained probability values were added to the structural features in Table 2 and the final features were trained with the XGBoost model. The performances of the model for the 5-fold are shown in Table 3 in the format mean ± standard deviation.

Table 3. 5- Fold results of the ensemble model approach (mean ±standard deviation

XGBoost with	Accuracy		F1 score	
	Segment Based	Machine Based	Segment Based	Machine Based
CNN Architecture	0.8334±0.017	0.8289±0.03	0.747±0.055	0.761±0.0505
EfficientNet B0	0.85± 0.026	0.83± 0.014	0.779±0.029	0.762±0.028

EfficientNet B7	0.8511±0.03	0.8452±0.017	0.777±0.031	0.783±0.029

Segment-based results show the results evaluated over each 3-second sound section, while machine-based results are the results in which the performance is measured by determining the average probability value and the final class in the machine group to which each sound belongs. It can be misleading to decide on the best model based on machine-based results, as not all sound sets are of equal length. Therefore, the EfficientNetB7+XGBoost ensemble model gave the best accuracy and F1 scores compared to other models.

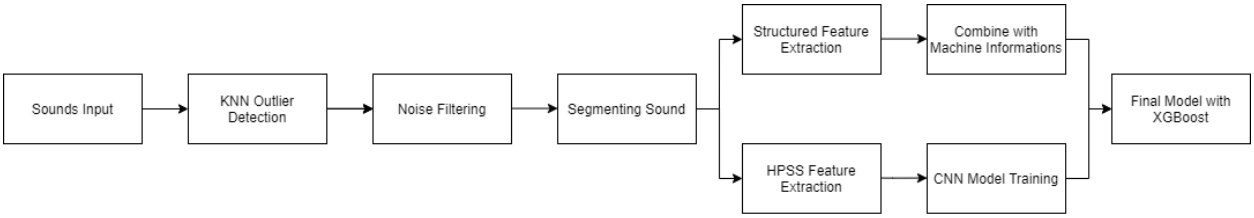


Fig 5. *Flowchart of our ensemble approach*

5. Conclusion

Machine sounds contain important information about the current status of the machine. The early detection of any malfunction is critical to identify the specialist personnel who will repair the machine.

Different data structures are used to determine the current status of construction machines. The traditional process was concluded with an innovative approach by making a fault prediction over the sound from a different perspective.

In this study, we propose an approach that uses a combination of CNN and gradient boosting. Machine sounds are first divided into 3-second sections, then the harmonic and percussive components of the meal spectrogram and the Mel spectrogram are extracted from these sections, and a 3-dimensional matrix is created. CNN, EfficientNetB0, and EfficientNetB7 models are trained with these matrices. Structural features were extracted with the same 3-second slices and these structural features were combined with the probability values obtained from the CNN models. An XGBoost algorithm is trained with the created attributes. According to the results obtained, the EfficientNetB7+XGBoost ensemble model provided the best performance.

References

[1] Jena DP, Panigrahi SN. Motorbike piston-bore fault identification from engine noise signature analysis. J Appl Acoust 2014; 76: 35-47.

[2] Fitzgerald, Derry "Harmonic/percussive separation using median filtering," 13th International Conference on Digital Audio Effects (DAFX10), Graz, Austria, 2010.

[3] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks," International Conference on Machine Learning, PMLR, 2019.

[4] Yang, Ruo-Yu, and Rahul Rai. "Machine auscultation: enabling machine diagnostics using convolutional neural networks and large-scale machine audio data," Advances in Manufacturing 7.2, 174-187, 2019.

[5] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system," Proceedings of the 22nd ACM signed international conference on knowledge discovery and data mining. 2016.