A Deep Learning Approach to the Acoustic Condition Monitoring of a Sintering Plant

Shahab Pasha[1,2], Christian Ritz[2], David Stirling[2],
Paul Zulli[1], David Pinson[3], Sheng Chew[3]
[1] *ARC Research Hub for Australian Steel Manufacturing*
[2] *School of Electrical, Computer and Telecommunication Engineering, University of Wollongong, NSW, Australia*
[3] *Coke and Ironmaking Technology, BlueScope Steel, Port Kembla, NSW, Australia*

***Abstract*— This paper proposes the use of deep learning classification for acoustic monitoring of an industrial process. Specifically, the application is to process sound recordings to detect when additional air leaks through gaps between grate bars lining the bottom of the sinter strand pallets, caused by thermal cycling, aging and deterioration. Detecting holes is not possible visually as the hole is usually small and covered with a granular bed of sinter/blend material. Acoustic signals from normal operation and periods of air leakage are fed into the basic supervised classification methods (SVM and J48) and the deep learning networks, to learn and distinguish the differences. Results suggest that the applied deep learning approach can effectively detect the acoustic emissions from holes time segments with a minimum 79% of accuracy.**

***Index Terms*—Acoustic monitoring, Deep learning, Supervised machine learning, Sintering plant**

I. INTRODUCTION

Acoustic monitoring has a wide range of applications from whale migration tracking [1], to intrusion detection [2] and health assessment [3] due to its non-intrusive nature and the limitations of the other types of signals such as video and thermal signals under particular circumstances. Industrial applications of acoustic monitoring also cover a wide range of machineries and environments, from agricultural cutting machinery [4], to corrosion and aging monitoring of valves and pipes [5]. It is said that acoustic condition monitoring is one of the most efficient strategies for identifying maintenance requirements in industry [6].

Continuous monitoring of industrial machinery and detecting, or predicting, associated faults and hazards, can significantly decrease maintenance and operational costs. Acoustic monitoring is used in the steel industry for applications such as temperature monitoring of slide gate plates, monitoring of the temperature inside a lining of a metallurgical vessel and Radio Frequency IDentification (RFID) tagging of slag ladles [7]. In this study, the application of acoustic monitoring within an iron ore sintering plant is proposed and evaluated using deep learning methods.

Iron ore sintering is a thermal agglomeration process, utilising a blend of fine materials (iron ores, fluxes, solid fuels, etc) to produce sinter of specified metallurgical quality and strength. Sinter is one of the main burden materials charges into a blast furnace.

The sinter blend is distributed onto a travelling grate made up of steel pallets (Figure 1), the bottom consisting of steel grate bars. The top surface of the granular bed is ignited and air drawn downwards, under suction. As the solid fuel combusts and the air pulls the reaction zone (or flame front) down through the bed, the energy generated partially melts the sinter blend, which in turn, helps to bond the constituent materials together. Under normal conditions, this air generates a stationary and relatively low frequency (2-5 kHz) noise. However, as the travelling pallets (and grate bars) undergo repeated heating and cooling, it is difficult to maintain the tight gap tolerances between bars. As gaps between bars increase, additional air is drawn through them, generating a slightly high-pitched whistle sound (6-8 kHz).

a)

b)

Figure 1: a) Sinter strand, with two spare pallets on the side b) Detail showing strand made up of individual pallets. Each pallet has two sets of two wheels on each side, identified by a unique number.

Detection of this condition (and additional air leakage) is crucial for process control and energy efficiency. As the grate bars are covered with material, they are not visually accessible; hence, this deformation monitoring is very challenging and image processing techniques are not applicable.

In this paper a deep learning approach is proposed and compared with the standard machine learning approaches, which are based on feature extraction, training and supervised classification. Raw, short-term, time-frequency signals are fed into a Recurrent Neural Network (RNN) [8] as the applied deep learning tool. Support Vector Machine (SVM) [9] and J48 [10] pruned decision tree classifiers are also applied to the derived features as basic machine learning methods.

The remainder of this paper is organised as follows. Section II describes the main blocks to a supervised acoustic monitoring system. Section III and Section IV are dedicated to the process of the feature extraction from the raw signals (where needed). The classification results are presented in Section V and the paper is concluded in Section VI.

II. Supervised Classification of Sinter Plant Noise

The classic process of acoustic classification consists of three main stages: Data collection (sinter plant recordings); feature extraction; and supervised classification using a trained classifier [11]. To train the classifier, a hand-labelled database of recorded signals for each class (e.g. "W" for Whistle and "R" for Regular) is required. A fraction (e.g. 50%) of the data are used as the training set to produce the trained classifier as shown in Figure 2.

The recorded signal $x(n)$, is mathematically modeled as

$$x(n) = a(n) + w(n) + v(n), \tag{1}$$

where n denotes the discrete time index, $a(n)$ is the stationary air noise and $w(n)$ and $v(n)$ represent the air-leakage (whistle) and the background noise respectively. The time-frequency domain representation of the signal is

$$x(n,k) = a(n,k) + w(n,k) + v(n,k), \tag{2}$$

where n, k are the time and frequency indices, $a(n,k)$ is the time-frequency representation of the stationary air noise, $w(n,k)$ is the occasionally occurring whistle noise and $v(n,k)$ is the background noise consisting of the mechanical noise and other process related sounds.

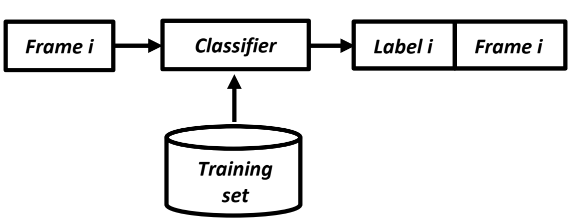


Figure 2: Supervised classification

a)

b)

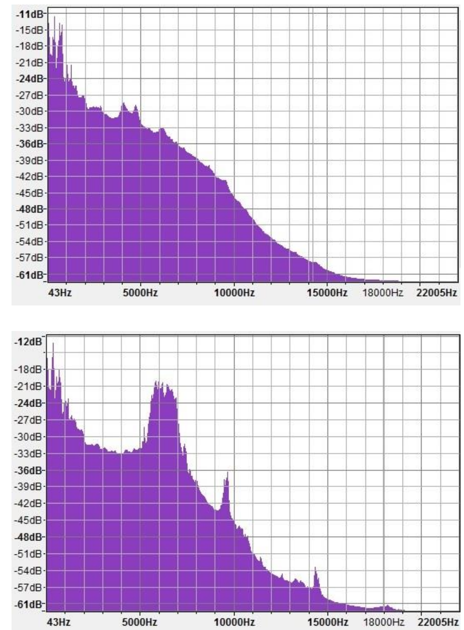


Figure 3: Preliminary signal analysis: a) Magnitude spectrum of the regular air noise (two minutes) b) Magnitude spectrum of the whistle noise (two minutes), NFFT=1024,f_s=44100 .

Here, the deep learning classifier is trained on time-frequency representations of the sintering plant noise $(x(n,k))$ and the extracted acoustic features of the recorded signal $(x(n))$. The (2) is obtained using a Fast Fourier Transform (FFT) applied to rectangular windowed time frames of length N_{FFT}.

As it was observed through preliminary experiments (Figure 3), the air leakage noise (the whistle) results in higher amplitude peaks in the magnitude spectrum of (2) around 6 to 8 kHz (Figure 3(b)) compared to the magnitude spectrum of regular air noise (Figure 3(a)). It can also be observed that the magnitude spectrum of the whistle noise generally has a number of additional peaks compared to the regular air noise. These differences indicate the potential to detect the presence of the whistle, based on classifying the spectra or derived features.

It was also observed that the process air and air leakage noises do not have constant absolute amplitudes; depending on the material laid on the pallets, temperature and the pallets conditions, the whistle and the regular air noise may change (in terms of the amplitude and the frequency). However this change is not sudden and happens slowly.

III. Discriminative Features of Sinter Plant Recordings

Based on the observations of Section II, two classification approaches are considered. First, using the ratio of the power corresponding to the whistle band (6 - 8 kHz) to the power of the full band signal (0 kHz- 22 kHz) as a discriminative feature [12] within machine learning-based classifiers.

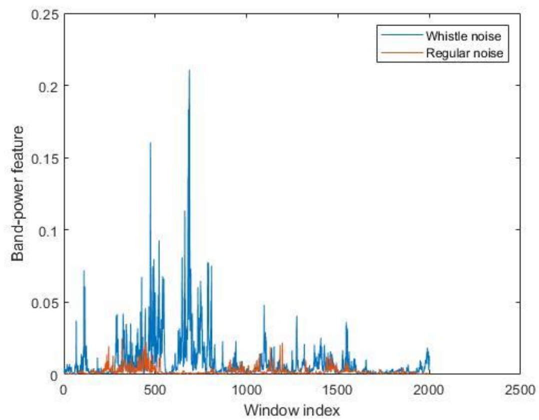


Figure 4: Band-power feature over time (2000 frames) of a whistle and a regular noise

Second, using the raw time-frequency spectrograms as input to the deep learning classifiers (RNN). In other words the deep learning classifier is applied to distinguish and learn the differences between the two classes of signals.

A. *Band-Power Ratios*

The band-power features (3) [12] are applied as the discriminative feature in order to detect the time-segments with unusual power distribution.

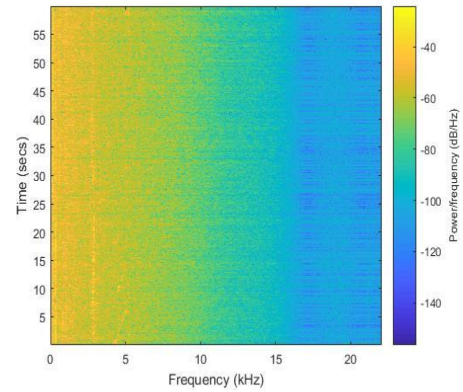
$$F = \frac{\sum_{l \in w_i} |X(n,l)|^2}{\sum_{k=0}^{N-1} |X(n,k)|^2}, \tag{3}$$

where w_i is the set of power spectrum samples belonging to the i-th band within the whistle frequency range. In this study, the target band (w_i) from (2) is 6 - 8 kHz, the total band width is 22 kHz and (3) is calculated once per time frame (1024 samples). Figure 4 shows the band-power feature extracted for one whistle and one regular recording. The band-power is calculated for every 1024 samples (one window length) at 44100 sampling rate. As shown, the power of the target band increases significantly (up to 0.2) when there is air leakage whereas in normal (regular) situations, the band-power of the target band (6 - 8 kHz) is a lower value (around 0.02).

B. *Time-Frequency Spectrograms*

Rather than deriving features for each time-frame, the spectrogram can be used as the input to a classifier, as used in other acoustic scene classification applications [13]. The spectrogram is formed from the sequence of frame-based magnitude spectra calculated in (2). Figure 5 shows the spectrogram derived for one minute of a whistle and non-whistle (regular) recording at the sinter plant.

a)



b)

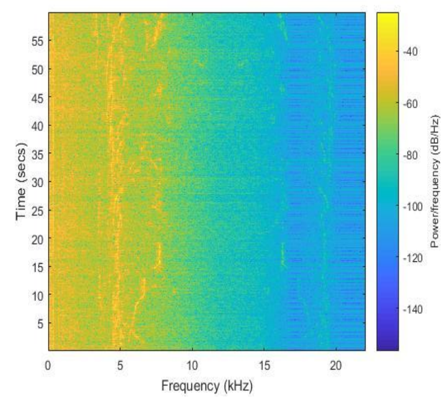


Figure 5: a) Spectrogram of one segment of the regular signal, b) Spectrogram of one segment of the whistle signal

In order to apply a deep learning method on the raw time-frequency signals, the FFT (4) of each window is calculated and the vector of N_{FFT} real values are applied as the feature for each frame.

$$X(k) = \sum_{n=0}^{N_{FFT}-1} x(n) e^{-i\omega_k n} \tag{4}$$

where $\omega = \frac{2\pi k}{N_{FFT}}$ for $k = 0,1,\ldots,N_{FFT}-1$. The feature vector for each frame Figure 5 is the magnitude of $\{X(0),\ldots,X(k),\ldots X(N_{FFT}-1)\}$ from (4). The FFT vectors are then averaged (5) and hand-labelled across time for each segment T_j (one minute) of the recording to obtain one vector. This decision making interval (one minute) is chosen based on the pallet movement speed and the monitoring requirements advised by the industrial experts familiar with the sintering process.

$$\bar{X}(j) = \sum_{T_j} \frac{X(k)}{T_j}. \tag{5}$$

IV. Applied Machine Learning Techniques

The following machine leaning techniques are applied to the extracted features and signals described in Section III. The goal is to obtain the classification results for each classifier-feature combination for the same dataset and compare the effectiveness of the methods and features.

A. *Support Vector Machine*

A Support Vector Machine (SVM) [9] is applied as the basic binary classifier. The problem is modeled as

$$\vec{G}.\vec{x} - b = \pm 1 \quad \text{for two classes,} \tag{6}$$

where $\vec{G}.\vec{x} - b = \pm 1$ represents two hyperplanes covering two classes and $\vec{G}$ is the weightings. The vector $\vec{x}$ is the feature vector from (3) and (5). It is noteworthy in the case of using the band-power as the feature that the two hyperplanes are actually two half-lines and the margin between the two classes is a point picked by the algorithm based on the data distribution (simply a threshold).

B. *J48 pruned decision tree*

J48 (also referred to as C4.5) [10] data mining algorithm uses the training set to build decision trees based on information entropy. A J48 tree is built similar to any other decision tree by finding the best splitting attribute however in order to address the outliers and overfitting issues a pruning process is added. Pruning is carried out from the leaves to the root. The J48 method is briefly explained in Figure 6.

C. *Deep learning for spectrogram classification*

Recurrent Neural Network (RNN) with a large number (more than 100) of hidden layers is applied as deep learning tool. The RNN considers the current input and the past state of the hidden layer for decision making (classification). In other words the hidden layer (state) at each time, t is a function of the current input and the past state.

$$h(t) = \emptyset(V \times i(t) + U \times h(t-1)). \tag{7}$$

where $h(t)$ is the current state of the hidden network, $\emptyset(.)$ is the logistic sigmoid function and describes the dynamic to adjust the significance of the current input and the past state [14] characteristics of the RNN, $x(t)$ the current input and $h(t-1)$ the previous hidden state. W and U are the applied weightings as illustrated in Figure 7 and Table 1.

J48 algorithm	
1.	*The potential information is calculated for every attribute*
2.	*The best attribute resulting in the highest gain in information is chosen for branching*
3.	*The tree is generalised by removing the specialised branches caused by the training set*

Figure 6: J48 algorithm

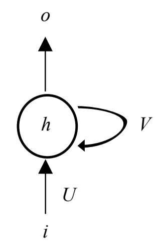


Figure 7: RNN recurrent hidden layer

Table 1
RNN network architecture

Quantity	***Value/Method***
Number of the hidden layers	*150*
Maximum number of training iteration	*1000*
Deep learning algorithm	*Levenberg-Marquardt*
Adaption learning function	*Gradient descent with momentum weight and bias*
Neural network transfer function	*Hyperbolic tangent sigmoid transfer function*

Table 2
Experimental setup

Symbol	***Quantity***	***Value/Method***
T	*Segment length*	*60 s*
f_s	*Recording Sampling frequency*	*44100*
N_{FFT}	*Number of the frequency bins*	*1024*
W	*Window length*	*1024 samples 23ms*
	Window type	*Rectangular*
N_{Ov}	*Overlap*	*50% 512 samples*

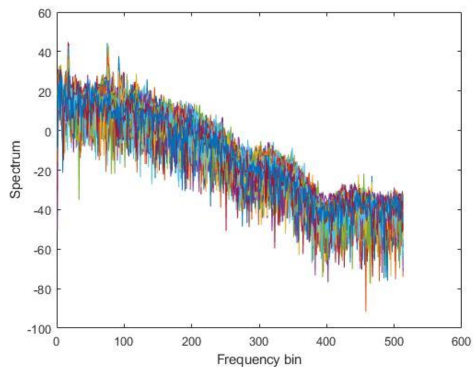
V. Experimental Studies

In this section the recorded signal, feature extraction, training and evaluation process are described. The classification results are presented for each classifier and feature. Deep learning method (RNN) is applied to learn the difference between the raw time-frequency signals and distinguish the whistle and regular segments.

A. *Experimental setup*

132 minutes (Figure 8), 341085 window lengths of the sintering plant noise is recorded by a 130A24 PCB water and dust resistant microphone [15]. The spectrograms and the band-power features calculated for each window (1024 samples

a)



b)

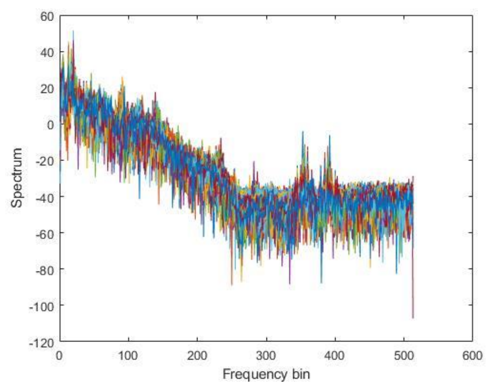


Figure 8: Spectrum data set for each class. a) 40 whistle segments b) 92 regular segments

which translates to 23ms at 44100 kHz sampling rate) are calculated and stored. As the noise is stationary and changes do not occur suddenly, the spectrograms and the band-power features are averaged for each segment (one minute worth of signal) and hand labelled. 50% (66 minutes) of the labelled segments are randomly chosen and used for training a J48 pruned decision tree, an SVM classifier and a deep learning RNN with 150 hidden layers (Table 1 and Table 2). The other 50% of the data are unseen by the classifiers and are applied for evaluation (25% for validation and 25% for testing). This process is repeated 10 times (in order to cancel the effect of the randomly chosen training and evaluation sets) and the averaged results are presented in this section.

B. *Classification results*

Having a confusion matrix as:

$$\begin{bmatrix} & \textit{Predicted as "R"} & \textit{Predicted as "W"} \\ \textit{True class "R"} & TN & FP \\ \textit{True class "W"} & FN & TP \end{bmatrix} \quad (8)$$

("R" for regular and "W" for whistle) the following classification evaluation measurements (9)-(14) are applied [16] where True Positive Rate (TPR) shows how accurately the whistle frames are detected and False Positive Rate (FPR) shows how often a "regular" frame is labelled as a "whistle" frame. Precision, Recall, Accuracy and the F-score indicate the overall success rate of the classifier in distinguishing between the "whistle" and "regular" frames.

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$FPR = \frac{FP}{FP + TN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Fscore = \frac{2.Precision.Recall}{precision + Recall} \quad (14)$$

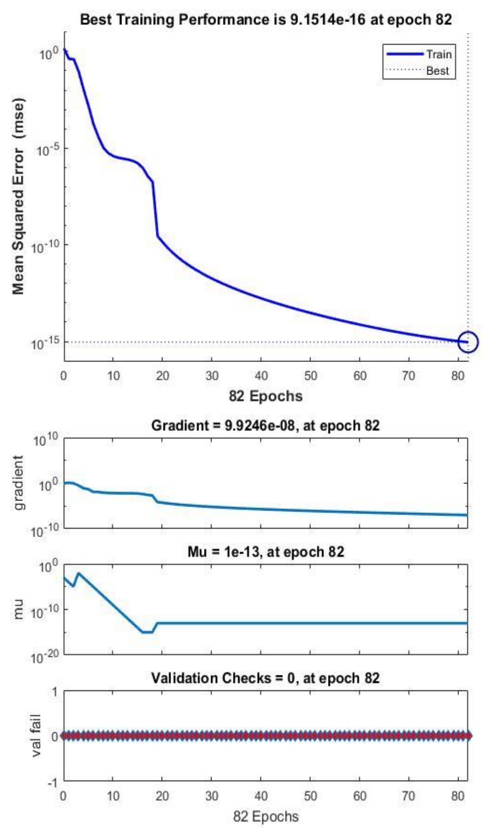


Figure 9: RNN training evaluation

Table 3
CLASSIFICATION ACCURACY-BAND-POWER FEATURE

Applied feature	Applied classifier	Classification accuracy
Band-power	RNN	**79.55%**
Band-power	J48	74.24%
Band-power	SVM	67.42%

Table 4
CLASSIFICATION ACCURACY –DEEP LEARNING

Applied feature	Applied classifier	Classification accuracy
Spectrogram	RNN	**81.80%**
Spectrogram	J48	75.00%
Spectrogram	SVM	73.67%

Table 5
CONFUSION MATRIX-J48 (BAND-POWER)

Class	Whistle	Regular
TP rate	0.79	0.71
FP rate	0.28	0.20
Precision	0.57	0.87
Recall	0.79	0.71
F-score	0.66	0.79
ROC area	0.54	0.84

Table 6
CONFUSION MATRIX-RNN (SPECTROGRAM)

Class	Whistle	Regular
TP rate	0.77	0.83
FP rate	0.22	0.16

Table 7
CONFUSION MATRIX-RNN (BAND-POWER)

Class	Whistle	Regular
TP rate	0.75	0.81
FP rate	0.25	0.19

Figure 9 shows the RNN training status and the mean square error after 82 epochs (iterations) where the cost function optimisation criteria is met. The classification results for the applied classifiers (Section IV) are presented in Table 3 and 4. It is shown that RNN outperforms the J48 and SVM classifiers where the classifiers are applied under the same training and evaluation policy. It is also shown that the RNN applied to the raw, time-frequency signals yields the highest level of accuracy (81.8%) (Table 4).

Table 5 shows the classification accuracy measurements for a J48 decision tree. It is shown the trained tree can effectively detect the whistle segments in 79% and the Regular segments in 71% of the times. Table 6 and 7 indicate how accurately the applied RNN is able to classify each class of the input data whereas Table 3 and

Table 4 show the overall accuracy across the two classes. Figure 10 is the RNN Receiver Operator Characteristic (ROC) for training, validation, testing and the overall ROC with the ROC area of 0.91.

Experimental studies of this research show that the classifiers provide stable, accurate air-leakage detection results. However the SVM classifier when applied to the raw time-frequency signals does not provide stable results and the accuracy changes from 63% to 80% depending on the training and the testing set.

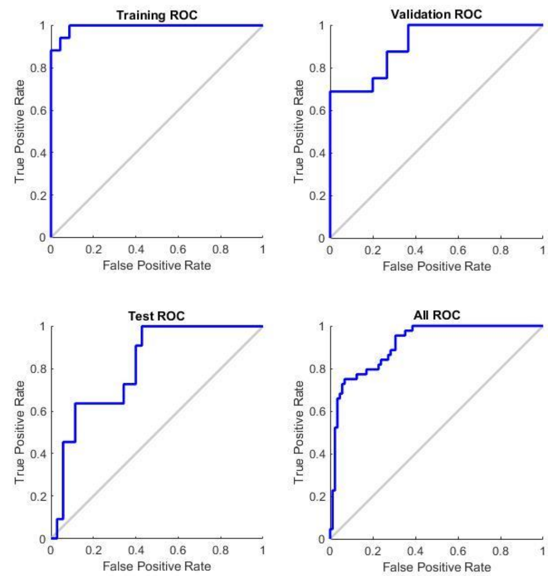


Figure 10: RNN ROC-Spectrogram

The spectrogram features outperform the band-power feature as they contain information from all the frequency bands whereas the band-power ratios only rely on the relative power of the whistle band. As it was shown in Figure 8 the first 100 (out of 513 bins) frequency bins and the last 100 bins of both classes are similar and the main difference is in the middle band frequency bins.

VI. CONCLUSION

Deep learning and basic supervised classifiers are successfully applied to the task of acoustic condition monitoring of the sintering plant. The proposed deep learning approach successfully detects the presence of irregular leakage noise (whistle) associated with pallet faults. The Results suggest that the band-power feature is an effective feature for baseline machine learning methods (SVM and J48) however the raw spectrogram pattern of the air-leakage noise is also learnt and recognized accurately by the deep learning method (81.8% accuracy). It is concluded that deep learning neural network method (RNN) outperforms the J48 pruned decision tree and SVM binary classifier. Future work will focus on applying unsupervised machine learning techniques to the problem of detecting the unusual condition at the sintering plant.

Acknowledgement

The authors gratefully acknowledge the funding from the Australian Research Council Industrial Transformation Research Hubs Scheme (Project Number IH130100017) and the in-kind contributions of BlueScope Steel personnel.

REFERENCES

[1] R. Davis, M. Baumgartner, A. Comeau, D. Cunningham, K. Davis, A. Furlong and H. Johnson, "Tracking whales on the Scotian Shelf using passive acoustic monitoring on ocean gliders," in *OCEANS 2016 MTS/IEEE*, Monterey, CA, 2016.

[2] C. Zieger, A. Brutti and P. Svaizer, "Acoustic Based Surveillance System for Intrusion Detection," in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, Genova, 2009.

[3] A. Ukil, S. Bandyopadhyay, C. Puri, A. Pal and A. Mukherjee, "Heartmate: automated integrated anomaly analysis for effective remote cardiac health management," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017.

[4] H. Siebald, "Real-Time Acoustic Monitoring of Cutting Blade Sharpness in Agricultural Machinery," *IEEE/ASME Transactions on Mechatronics,* vol. 22, no. 6, pp. 2411-2419, Dec. 2017.

[5] W. Lin, S. Wu, Y. Shi and F. Wang, "Research on non-intrusive acoustic monitoring for valve internal leakage," in *29th Chinese Control And Decision Conference (CCDC)*, Chongqing, 2017.

[6] E. Cano, J. Nowak and S. Grollmisch, "Exploring sound source separation for acoustic condition monitoring in industrial scenarios," in *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017.

[7] R. Fachberger and A. Erlacher, "Applications of wireless SAW sensing in the steel industry," *Procedia Engineering,* vol. 5, pp. 224-227, 2010.

[8] M. Kolbaek, D. Yu, Z. H. Tan and J. Jensen, "Multitalker Speech Separation With Utterance-Level Permutation Invariant Training of Deep Recurrent Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 25, no. 10, p. 1901, 2017.

[9] W. Lipo, Support Vector Machines: Theory and Applications, Springer, 2005.

[10] J. Bell, Machine Learning: Hands-On for Developers and Technical Professionals, Wiley, 2015.

[11] J. Ye, T. Kobayashi, M. Murakawa and T. Higuchi, "Robust acoustic feature extraction for sound classification based on noise reduction," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.

[12] A. Riid, S. Astapov, J. Preden and T. Aruväli, "Industrial Process Monitoring by Multi-Channel Acoustic Signal Analysis," in *14th Biennial Baltic Electronic Conference (BEC)*, Tallinn, 2014.

[13] TUT, "DCASE challenge," Tampere University of Technology, 2017. [Online]. Available: http://www.cs.tut.fi/sgn/arg/dcase2017/challenge/task-acoustic-scene-classification-results. [Accessed March 2018].

[14] Y. Zhao, X. Jin and X. Hu, "Recurrent convolutional neural network for speech processing," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017.

[15] M. S. CORPORATION, "PCB Piezotronics," 2017. [Online]. Available: http://www.pcb.com/contentstore/MktgContent/LinkedDocuments/Acoustics/TM-AC-130A24_LowRes.pdf. [Accessed 14 may 2018].

[16] D. D. Gutierrez, Machine Learning and Data Science: An Introduction to Statistical Learning, NJ, USA: Technics publication, 2015.