# Early Failure Detection of Belt Conveyor Idlers by Means of Ultrasonic Sensing

Daniel R. Ericeira*, Filipe Rocha†‡, Andrea G. C. Bianchi*, Gustavo Pessin*†
*Federal University of Ouro Preto, Ouro Preto – MG, Brazil
†Instituto Tecnológico Vale, Ouro Preto – MG, Brazil
‡Federal University of Rio de Janeiro, Rio de Janeiro – RJ, Brazil
daniel.ericeira@gmail.com, andrea@iceb.ufop.br, {filipe.rocha, gustavo.pessin}@itv.org

***Abstract*—Belt Conveyors are the main class of machinery that compose the logistics of a port terminal. The rolling components of the conveyor may fail mainly due to damaged idlers, which may cause a severe industrial breakdown. Nowadays, the equipment protection is done by a set of sensors that indicate an already existing abnormality or by personal inspection, applying practical experience in the search of visual, sound, or temperature signatures of imminent failure. Aiming to upgrade from the current corrective system to the predictive domain, a model for early failure detection on the conveyor's idlers is proposed. Ultrasound recordings were conducted on idlers that did not present any perceptible abnormalities, labeled as non-defective, and on idlers that displayed typical failure noise, labeled as defectives. The dataset collected was used for the training and testing of Random Forest and Multilayer Perceptron machine learning algorithms. Four types of experiments were devised to test the methodology, two of them using time-domain data, and two of them using frequency domain data, with different statistical attributes. The results achieved in various classification experiments showed that there is a distinctive pattern on the ultrasound spectrum that differs non-defective from defective idlers, as pre-evaluated by traditional methods of human inspection. In the best case, the experiment that used a moving average on the frequency domain data presented an average of 83.68% of correctly classified idlers, obtaining as best result accuracy of 89.47%.**

## I. Introduction

Belt Conveyors (BC) systems are the leading logistics machinery on a mining port terminal. This equipment continuously handles the unloaded bulk material for stacking in the stockyards and subsequent reclaiming, transportation, and shipping. Port production highly depends on the proper operation of such equipment.

The typical Belt Conveyor structure comprises a tensioned rubber belt directly supported by rollers and pulleys typically driven by electric motors. A frame holds one central and two angled rollers that are commonly installed each 1 m of conveyor section. A counterweight system stretches the belt and absorbs operational load variation, keeping the belt continuously in contact with the rollers. After passing by an end-pulley, the top belt section returns to the initial position by the structure inferior part, being supported by large rollers. Fig. 1(a) presents the scheme of a common BC structure, while Fig. 1(b) displays its cross-section.

The major rolling parts failure comes from damaged bearing, lack of lubricant fluid, dirt incursion, or manufacturing faults [3], [4]. It is estimated that 43% of bearing failures

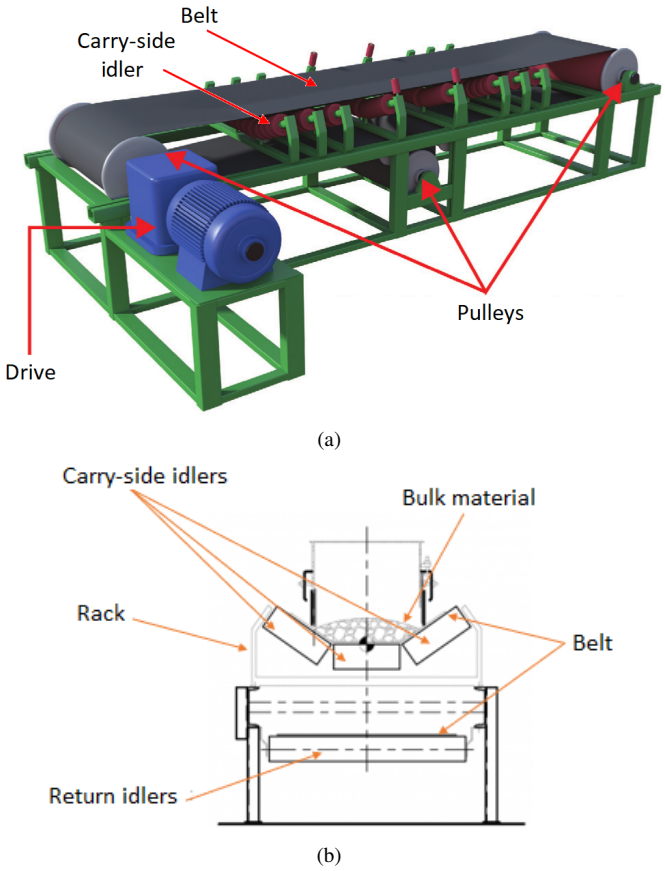



Fig. 1. (a) Simplified scheme of a Belt Conveyor, adapted from [1]. (b) Conveyor cross section, with carry-side and return idlers, adapted from [2].

are due to moisture or other contaminants caused by faulty sealings [5]. Failures on this component commonly block the roller, which leads primarily to an over power consumption due to motion resistance [6]. Furthermore, an obstructed roll may heat up given the belt friction, reaching up to 400°C [6]. Such a situation presents high damage potential as the heat may cause a fire in the surrounding structures.

Techniques that use the acoustic signal for machinery failure detection in the industry are successfully applied in [7], [8]. This sensing method emulates specialized human hearing, providing continuous monitoring, and automatic diagnostics

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

possibility. Acoustic Emission (AE) techniques base on direct sound capture and are complementary to traditional vibration and temperature measurements. Data collection can be performed by local sensors with communication capabilities, or by the self-contained microphones passing by the conveyor side.

Aiming to improve belt conveyor inspection techniques, we address the rollers' structural health monitoring problem using ultrasound signals. We recorded ultrasound data from rollers of an active conveyor at the Ponta da Madeira Mining Port using an Ultraprobe 10000 sensor, labeling them as *non-defective idler* (NDI) and *defective idler* (DI). BC specialists provided the roll health status considering its visual and operational aspects, abnormal vibration, eccentric rotation, noise emission, sealing lack, oxidation, and elongated roll extremities. We propose and evaluate different data structures (features) used as inputs of machine learning algorithms. The ultrasound sensing along the Belt Conveyor, combined with Artificial Intelligence techniques and signal pattern recognition is a potential method to enable early failure detection of belt conveyor idlers.

The remainder of this work is as follows: Section II reviews the related literature, presenting smart structural health monitoring and failure detection concepts. Section III briefly presents the employed machine learning techniques. Section IV describes the proposed methodology, detailing the data collection and the developed feature extraction types. Section V presents results and discussions. We finish the document presenting the conclusions and pointing potential future work.

## II. Related Work

Several works cope with intelligent inspection and failure detection. The most common signals type are vibration, acoustic emission, temperature, ultrasound and mechanical stress analysis. We focused on related works dealing with industrial equipment monitoring and diagnostics of rotating parts using sound. Prior works are grouped in two categories: (i) theoretical, proposing methods for using intelligent monitoring in large mechanical equipment; (ii) restricted to laboratories, performing tests under ideal conditions. One of the contributions of this paper is to apply the theoretical ground of prior research on working equipment on an industrial plant. Thus, we evaluate the applicability of such methods in real-world conditions.

Arredondo *et al.* [9] developed a methodology for fault detection in induction motors using sound and vibration signals. The failure types analyzed were of mechanical nature, like bearing failure, mechanical unbalance and broken rotor bars. Microphones recorded the sound emitted by running engines. Faulty motors presented high-intensity sound in specific frequencies, which they suggest as specific failure indication. Healthy motors did not present intensity peak at similar frequencies, suggesting that failure detection via sound emissions is viable. Marquez and collaborators [10] presented a general state-of-the-art review of signal processing techniques and methods applied in wind turbines condition monitoring. Acoustic emission is one of the most promising signals noted in the authors' research for fault diagnosis in bearings, gears, rotors, and blades.

Kim *et al.* [11] compared ultrasound and vibration for health monitoring of bearings. They highlights that one ultrasound advantage is it adequate detection related to signal-to-noise relation. A laboratory test rig allowed to capture signals from good and faulty bearing, both in time and frequency domains, using different statistical methods for evaluation. Results indicated that faulty bearings present distinctive ultrasound characteristics in relation to non-faulty ones, and that ultrasound is more effective in failure detection than vibration signals, suggesting its potential use for automatic diagnosis.

Tchakoua *et al.* [12] present a review of health monitoring of wind turbines. Acoustic emission is considered suitable for failure detection in gearboxes, bearings, shaft, and blades. They point out that as advantages, ultrasound measurements have a larger frequency range and relatively higher signal-to-noise ratio. The main drawbacks are the need to be close to the sound source – ultrasounds waves have a higher dissipation rate on thin air when compared to sound waves – and only a few types of fault are present in the high-frequency range.

Kedadouche, Thomas and Tahan [13] applied Empirical Mode Decomposition for early failure detection in mechanical components, like bearings and gears. This technique decomposes the original time series signal into intrinsic mode functions. Statistical descriptors were calculated for each signal decomposition (rms, crest factor, kurtosis e skewness) and their efficiency in early failure detection was compared. The authors concluded that the decomposed signal is more sensitive to bearing failures than the original non-decomposed signal. Experiments performed on a test bench detected a damaged bearing of 40 microns in size, with further tests planned to evaluate several defects of different sizes to explore the full potential of ultrasound condition monitoring.

The related work researched showed that sound data emitted from mechanical equipment is an efficient way of detecting abnormalities and diagnosing their operational condition. They also may provide meaningful information that can feed an automatic fault detection system. Moreover, ultrasound signal is less susceptible to noise and can provide more predictive insights.

## III. Machine Learning Overview

To solve increasingly complex problems with a large amount of data, recent computer algorithms developments focus on biological systems and natural intelligence emulation. According to Engelbretch [14], such systems form part of the Artificial Intelligence (AI) field, composed by logic, deductive reasoning, expert systems, case-based reasoning, and machine learning. According to AI definitions gathered by Russell and Novig [15], they vary along two main dimensions: (i) thought processes and (ii) behavior processes. Some definitions measure success in terms of human performance and others measure success as an ideal concept of intelligence

(rationality). This provides four possible goals to pursue in AI methods: systems that think like humans; systems that think rationally; systems that act like humans; and systems that act rationally. The human-centered approach is an empirical science, involving hypothesis and experimental confirmation, while the rationalist approach involves mathematics and engineering.

Machine Learning (ML) is a subset of AI that has self-learning as a key feature. It applies statistical modeling to detect patterns and enhance performance based on data. Mitchell [16] defined: *"a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"*. In ML, data is commonly divided into the training and the testing sets. The training dataset provides information for fitting the model. The model thus can be tested with a non-seen dataset (test set). The resulting model can then perform pattern recognition, classification tasks and decision making tasks [17], [18]. We used a supervised learning method in this work. In this branch, one feeds the model with data with various features and the output data correct label. The algorithm then identifies patterns and creates a model that tries to generalize the obtained rules with new data.

Intelligent monitoring applications can also be found in other domains, like in aviation [19]. For transportation systems, Torres and colleagues [20] present an evaluation of machine learning models to avoid texting while driving. Fernandes and colleagues [21] present an ensemble of convolutional neural networks to deal with unbalanced datasets in an inspection directed to train wagons. Intelligent industrial monitoring can also be performed by mobile devices, like robots or drones. Garcia and colleagues [22] present the robot ROSI, a robotic device for belt conveyor inspection. Merriaux et al. [23] introduce the robot VIKING, aiming to perform industrial inspections autonomously. Tanaka and collaborators [24] present a robotic device for inspection, but they deploy an articulated mobile robot as a snake.

Random Forests (RF) are algorithms that create a decision tree combination. A decision tree is a set of decision-making steps that has two possible outcomes: true or false. As the questions are answered, the algorithm goes through the decision tree towards a classification or prevision. The trees are combined with each other to achieve better performance. RF prevents overfitting and efficient with a large amount of data [25], [26]. During the training stage, a supervised model learns to relate data (features) desired results for the algorithm to predict (target). The decision trees learn to calculate the best questions to use the features that were fed to them, to perform the best possible predictions. RF algorithms form a combination of decisions made by several trees, each tree of the forest initiating a random set of features and using a random set of data for the training stage, enhancing diversity with the best prevision rates achieved by all the randomly assembled trees [27].

Artificial Neural Networks (ANN) comprise several activation nodes (neurons) with different weights grouped in (at least) three layers: the input layer, the hidden layer(s), and the output layer. The neurons inter-connections are inspired by brain functioning. As mentioned by Boutaba *et al.* [28] and Zhang *et al.* [29], the Multi Layer Perceptron (MLP) method exhibit features such as the ability to learn complex patterns of data and generalize learned information. The MLP learning process is subdivided in three parts: stimulation by example; iterative weights update to minimize output error; and response in a new way as result of changes previously occurred [30], [31].

Machine learning is an AI field used on classification tasks, which are an integral part of this work. Neural networks, such as MLP, commonly learn patterns using complex data. On its turn, RF are more adequate for larger datasets, presenting better generalization and capacity to identify more important variables for the classification. We used classification algorithms to learn patterns in the ultrasound recordings of healthy and defective rollers and generalize to non-seen components.

## IV. Methodology

This section describes the proposed methodology. The steps followed are (1) ultrasound recording of operational rollers on industrial site, (2) pre-processing of the audio files in time domain and frequency domains (using the Fast Fourier Transform – FFT algorithm), (3) data samples grouping, (4) extraction of features, (5) sample labeling as non-defective or defective idler, (6) training and evaluation of machine learning algorithms (Random Forest and Multilayer Perceptron) using the dataset, test and performance evaluation of the classifier by labeling of new data. We represent the sound signal using a waveform as a function of time. This waveform represents the intensity of the sound at each time instant. The Fourier Transform (FT) decomposes time waveforms as a sum of sinusoids of different frequencies.

FFT is an algorithm that applies the Fourier Transform as a discrete domain. Discrete Fourier Transforms (DFT) are useful as they reveal periodicities in the input data as well as the relative intensity of the periodic components. FFT is an efficient algorithm for DFT computing. The number of computations required in performing the DFT was dramatically reduced by the FFT algorithm (the number of computations required for N points is reduced from the order of $N^2$ to $NlogN$ [32]).

### A. Data collectionn

We recorded ultrasound from rollers during plant operation using an acoustic detector, used for predictive inspection in the Ponta da Madeira Mining Port located at São Luís, Maranhão, Brazil (2º33'16.6"S, 44º21'24.4"W). The ultrasonic sensor is the Ultraprobe 10000 (as shown in Fig. 2(a) from UE Systems Inc., which is used in the terminal for monitoring the electrical power transmission grid. This sensor captures sound in the lower ultrasonic spectrum (from 20kHz to 100kHz) and has a data recording software that can export the data as waveform audio format (wave) files. The Ultraprobe uses a signal processing technique, known as *Heterodyning*, to

(a) Ultrasonic sensor Ultraprobe 10000 (UE Systems Inc.)..

(b) Ultrasound recording of a conveyor idler in the MTPM.

(c) TR-315K-09, one of the BC accessed for ultrasound data acquisition.

Fig. 2. Environment of the data acquisition.

map the high-frequency acoustic signal to the human-hearing range. The experienced human operator listens to patters in the transduced signal in search of defective idlers.

The recording was made by positioning the sensor close to the idler bearings, as recommended by the Ultraprobe manufacturer, during the conveyor operation (as showed on Fig. 2(b). The equipment noise sensitivity was set to minimize the external noise interference of other nearby equipment. Fig. 2(c) exhibits the Belt Conveyor structure in the TMPM port, accessed for ultrasound data acquisition. The goal was to create an ultrasound recordings dataset of idlers that did not exhibit any functioning abnormalities noticeable by the human senses (labeled as non-defective idlers – NDI), and recordings of idlers that exhibit noticeable abnormalities (through visual and auditive inspection, labeled as defective idlers – DI). The patterns perceived in these two conditions were used to train a set of classifier algorithms that had their performance further evaluated.

### *B. Pre-processing of the audio files*

We recorded ten time-series in non-defective idlers and ten recordings for the defective ones. Each recording lasted for approximately 20 seconds. For each recording, two basic pre-processing steps were performed: (1) data acquisition in the time domain and (2) transformation to the frequency domain. For the *time domain* data acquisition, each file was transformed as a sound amplitude data vector normalized between -1 and 1. The sampling rate was 16000 values for each audio recording second. Later, for each experiment, these vectors were divided into smaller parts to increase the number of samples.

A typical time-domain curve for a non-defective idler is shown in Fig. 3(a). Each second in the graph contains 16000 amplitude values normalized between -1 and 1. A curve displaying a defective idler pattern is shown in Fig. 3(b). For the *frequency domain* data transformation, on each FFT file, we applied with 4096 frequency range sections. Later, these frequency values would be divided in smaller intervals to form the testing dataset samples. A typical frequency domain curve for a non-defective idler is shown in Fig. 3(c). A curve displaying a defective idler pattern is shown in Fig. 3(d).

### *C. Investigated models of feature extraction*

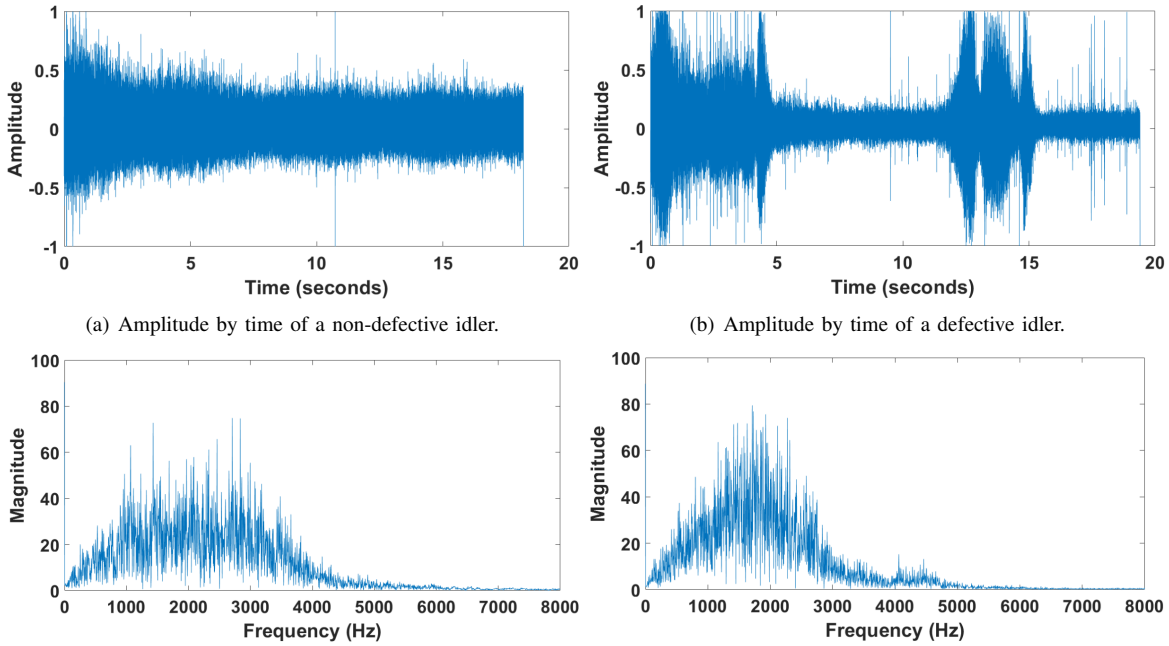
We devised four different feature extraction models with different sampling and characteristics. The proposed models are as follows:

**Experiment 1**: Half second sampling attribute vector in time domain; In E1, for each half second sample from the audio file, we computed the following statistical moments: mean, median and standard deviation of 8000 sound amplitude values. Along with these three values, each vector was labeled as NDI or DI regarding the type of idler. Thus, a dataset was assembled with 378 half-second samples labeled as NDI and 390 half-second samples labeled as DI. Each half-second resulted in three attribute and one label vector: mean, median and standard deviation and the label NDI or DI.

**Experiment 2**: One-second sampling attribute vector in time domain; In E2, for each one-second sample from an audio file, the following statistical moments were calculated: mean, median and standard deviation of every quarter of a second. A vector was thus assembled with 12 sound amplitude values (four mean values, four median values and four standard deviation values). Along with these twelve values, each vector was labeled as NDI or DI regarding the type of idler. Each second resulted in a twelve attribute and one label vector: 4 mean values, 4 median values and 4 standard deviation values and the label NDI or DI.

**Experiment 3**: FFT of a complete audio file divided into 40 parts; In E3, for each audio file, we computed the FFT with 4096 frequency ranges, spanning from 0 to 8000Hz. Each FFT was divided in 40 parts, composed by 102 frequency values each. For each part, the following statistical values were calculated: mean, median and standard deviation. Thus, a dataset was assembled with 400 samples labeled as NDI and 400 samples labeled as DI. Each one of the 40 parts of a FFT resulted in a three attribute and one label vector: mean, median and standard deviation and the label NDI or DI.

**Experiment 4**: Moving mean of the FFT of a 5 second sample; In E4, for each 5 second recording the FFT was computed, with 4096 frequency values ranging from 0 to 8000Hz. Each FFT was divided in 40 parts, composed by 102 frequency values each. For each part, the mean frequency




(a) Amplitude by time of a non-defective idler.

(b) Amplitude by time of a defective idler.

(c) Magnitude by frequency the non-defective idler shown in (a).

(d) Magnitude by frequency the defective idler shown in (b).

Fig. 3. Data in time domain and frequency domain of defective and non-defective idlers.

value was calculated. This generated a 40-attribute vector representing the moving mean for each 5 second sample. Thus, a dataset was assembled with 40 samples labeled as NDI and 40 samples labeled as DI.

### *D. Training and evaluation test of the classifier*

We used the total number of samples classified as NDI and DI in the training step of a RF and MLP machine learning algorithms to achieve diversity in the classification task and enable means of comparison between then. The complete dataset was deployed to run the RF and MLP classifiers using the machine learning collection Waikato Environment for Knowledge Analysis (WEKA) [33]. The experiments took place with ten-fold cross-validation.

Each experiment initiated with 10 different seed values (starting conditions) for comparison effects. Each experiment carried out the following RF and MLP configurations: RF with 10 trees (RF10), RF with 50 trees (RF50), RF with 100 trees (RF100), MLP with one 5-neuron layer (MLP5), MLP with one 10-neuron layer (MLP10), MLP with two 5-neuron layers each (MLP5x5) and MLP with two 10-neuron layers each (MLP10x10). Boxplots were assembled with the correct classification (hit) rate achieved by each experiment.

## V. Results

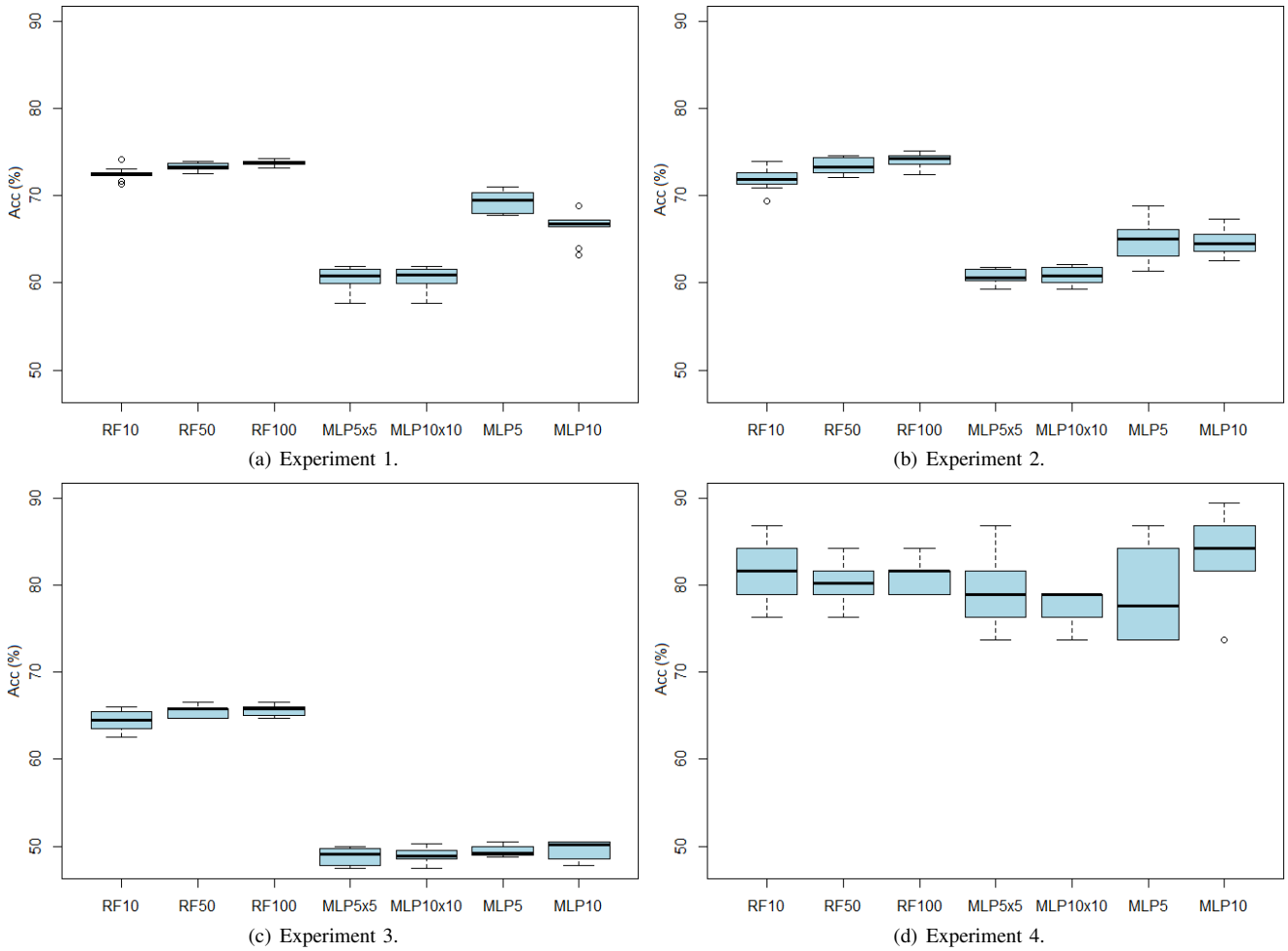
### *A. Investigation with ML using time domain data*

Experiments 1 and 2 take into account time-domain data. Experiments 3 and 4 take into account frequency domain data, by means of the FFT algorithm applied on the time domain data. For each different experiment (different input features) we run RFs in the following configurations: 10, 50 e 100 trees, and MLPs with the following configurations: one layer of 5 and 10 neurons, and two layers with 5 and 10 neurons each. Each experiment was run 10 times with different seed values. Ten-fold cross-validation was applied.

The first experiment (E1) was performed employing half-second samples extracted from the 20 idler recordings. The best result for the RF was achieved with the RF with 100 trees; the RF100 algorithm presented 74.21% of correctly classified instances. Of the 378 non-defective idlers, 295 were correctly labeled while 83 were mistakenly classified as defective. Of the 390 defective idlers, 275 were correctly labeled while 115 were mistakenly classified as non-defective. In E1, the MLP presented inferior classification results in comparison with the RF algorithms. Results are presented in Fig. 4(a).

We carried out a statistical evaluation upon the results presented in Fig. 4(a) aiming to evaluate the behavior of the methods. We employ the Shapiro-Wilk normality test to verify the adequacy of the results to parametric or non-parametric distributions. For all sets in E1, results presented p-values larger than 0.01. In this case, we can consider that, with 99% of confidence, the results fit normal distributions and a t-test (Welch Two Sample) can be employed. In E1, the best result was found using RF100. The comparison of the RF100 with the others showed that RF100 is equivalent to RF50 (Welch Two Sample t-test p-value 0.018). In this case, we can consider that, with 99% of confidence, the results of RF100 and RF50 are equivalent in E1. All other comparisons showed p-values lower than 0.01.

The second experiment (E2) was performed employing one-second samples with a 12-attribute vector with means, medians and standard deviations extracted from the 20 recordings.




(a) Experiment 1.

(b) Experiment 2.

(c) Experiment 3.

(d) Experiment 4.

Fig. 4. Accuracy for the different ML algorithms and different feature extraction models (Experiments E1 to E4).

The best result was achieved by the RF100 algorithm, which produced 75.12% of correctly classified instances when applied with cross-validation. By observing Fig. 4(b) one can notice that MLP presented a significantly inferior output. The two-layer MLP neural networks (MLP5x5 and MLP10x10) presented, in the worsts cases, 59.29% of correctly classified instances.

We also carried out a statistical evaluation upon the results presented in E2 (Fig. 4(b)) aiming to evaluate the behavior of the methods. For all sets in E2, results presented p-values larger than 0.01. In this case, we can consider that, with 99% of confidence, the results fit normal distributions and a t-test (Welch Two Sample) can be employed. In E2, the best result was found also using RF100. The comparison of the RF100 with the others showed that RF100 is equivalent to RF50 (Welch Two Sample t-test p-value 0.209). In this case, we can consider that, with 99% of confidence, the results of RF100 and RF50 are equivalent in E2. All other comparisons showed p-values lower than 0.01.

### *B. Investigation with ML using frequency domain data*

Experiment 3 employed FFT with a 3-attribute vector. It was performed on the samples extracted from the FFT applied in each of the 20 recordings divided into 40 parts. This way of sampling presented the worst results amongst the 4 experiments, as shown in Fig. 4(c). The RF algorithm has shown inferior results in relation to Experiments 1, 2 and 4, with 66.50% as the best accuracy in the 50 and 100 tree configurations, while the MLP presented as a worst result a 47.50% rate in the two-layer configurations.

We also carried out a statistical evaluation upon the results presented in E3 (Fig. 4(c)) aiming to evaluate the behavior of the methods. For all sets in E3, results presented p-values larger than 0.01. In this case, we can consider that, with 99% of confidence, the results fit normal distributions and a t-test (Welch Two Sample) can be employed. In E3, the best result was found also using RF100. The comparison of the RF100 with the others showed that RF100 is equivalent to RF50 (Welch Two Sample t-test p-value 0.730). In this case, we can consider that, with 99% of confidence, the results of RF100 and RF50 are equivalent in E3. All other comparisons showed

p-values lower than 0.01.

The way data was treated in E3 presented the worst results amongst the 4 experiments, as shown in Fig. 4. Also, we can see that, for E1 to E3, best results were found using RF with 100 trees, and the results are equivalent to the use of 50 trees.

Experiment 4 was performed with samples extracted from the FFT applied in each 5 seconds of the 20 recordings divided in 40 parts. This way of sampling presented the best results amongst the 4 experiments, as shown in Fig. 4(d). The best result was achieved by the MLP10, which in one case attained 89.47% of correctly classified instances.

We also carried out a statistical evaluation upon the results presented in E4 (Fig. 4(d)) aiming to evaluate the behavior of the methods. For all sets in E4, results presented p-values larger than 0.01. In this case, we can consider that, with 99% of confidence, the results fit normal distributions and a t-test (Welch Two Sample) can be employed. The only exception is the set MLP10x10 that showed a p-value lower than 0.000 and should be considered as a non-parametric distribution, in this case, a comparison using the Wilcoxon rank-sum test should be employed.

Best results among all were found in E4. In this experiment, the best result was found with the MLP10. We compared its results with other sets (RF10, RF50, RF100, MLP5, MLP5x5, MLP10x10). The comparison showed that all sets can be considered equivalent to 99% of confidence. The only exception is the comparisons of MLP10 with MLP10x10 that showed a p-value of 0.002 (Wilcoxon rank-sum test). Hence, both RF10, RF50, RF100, MLP5, MLP5x5, and MLP10 can be considered as equivalent for E4. The only MLP10x10 is not equivalent and presented the worst results, it might have happened due to overfitting.

Results of Experiment 4 emphasizes that, for the acquired data, both the modeling, the pre-processing methods and attribute definition are substantial.

The results achieved by this work were satisfactory in detecting the abnormalities previously perceived by human inspection. Related work in the field of failure detection by acoustic sensing was applied under controlled laboratory conditions with known defects; therefore, this work contributes to the application of the acoustic methodology in operational equipment on an industrial site.

## VI. Conclusion and Future Work

This work presented a methodology to early failure detection on belt conveyor idlers. Four types of experiments were devised (with time domain and frequency domain data with different attributes vectors) in order to evaluate which pre-processing and feature extraction methods are more efficient in achieving better classification results.

The results showed that the performance of the detection depends on the different data structures we employ as input of the classifier. Experiments 1 and 2 used time-domain data. The best case for Experiment 1 presented an accuracy of 74.22%. The best case for Experiment 2 showed an accuracy of 75.13%. The worst general results were achieved by Experiment 3, which employs the same attribute vector as Experiment 1 but using frequency domain data. The accuracy for the best case was 66.50%, with an average of 65.65%. The best general results were achieved by Experiment 4, which deploys a more detailed attribute vector in the frequency domain using a moving average with 40 values. For Experiment 4, most learning methods surpassed the accuracy of 80%. The best model for this Experiment showed an average accuracy of 83.68%, with the best case presenting accuracy of 89.47%.

Although this methodology presented satisfactory results, future work might explore even further the pre-processing techniques with other statistical moments, other ways of organizing attributes and other pattern recognition algorithms. A dataset with a long-term recording samples could also be assembled for better training of the model. Some future possibilities were envisioned during the development of this research, including (1) use of the intelligent system to identify specific failures on conveyor idlers; (2) long time monitoring to develop life expectancy prediction methods; (3) electronic and mechanics studies to adapt the system on a Belt Conveyor robot-inspector.

## References


[1] T. Schools, "Condition monitoring of critical mining conveyors," *Engineering & Mining Journal*, vol. 958948, pp. 50–4, 2015.

[2] "Conveyor belt design engineering," http://www.ingeniumdesign.us/services/conveyor-belt-design-engineering/, accessed: 2020-01.

[3] G. Lodewijks, "Strategies for automated maintenance of belt conveyor systems." *Bulk Solids Handling*, vol. 24, no. 1, pp. 16–22, 2004.

[4] C. Wheeler and D. Ausling, "Evolutionary belt conveyor design," in *The Int. Materials Handling Conf. and Exhibition-Beltcon*, vol. 14, 2007.

[5] Flexco, "What affects conveyor roller life?" *Insights*, 2011. [Online]. Available: http://documentlibrary.flexco.com/X2640_enAU_2525_INSCCTlife_0813.pdf

[6] R. Zimroz and R. Król, "Failure analysis of belt conveyor systems for condition monitoring purposes," *Mining Science*, vol. 128, no. 36, p. 255, 2009.

[7] M. Bowkett, K. Thanapalan, and J. Williams, "Review and analysis of failure detection methods of composites materials systems," in *2016 22nd International Conference on Automation and Computing (ICAC)*. IEEE, 2016, pp. 138–143.

[8] S. Grollmisch, J. Abe$\beta$er, J. Liebetrau, and H. Lukashevich, "Sounding industry: Challenges and datasets for industrial sound analysis," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019.

[9] P. A. Delgado-Arredondo, D. Morinigo-Sotelo, R. A. Osornio-Rios, J. G. Avina-Cervantes, H. Rostro-Gonzalez, and R. de Jesus Romero-Troncoso, "Methodology for fault detection in induction motors via sound and vibration signals," *Mechanical Systems and Signal Processing*, vol. 83, pp. 568–589, 2017.

[10] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, pp. 169–178, 2012.

[11] E. Y. Kim, A. C. Tan, J. Mathew, and B.-S. Yang, "Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements," *Australian Journal of Mechanical Engineering*, vol. 5, no. 2, pp. 177–189, 2008.

[12] P. Tchakoua, R. Wamkeue, T. A. Tameghe, and G. Ekemb, "A review of concepts and methods for wind turbines condition monitoring," in *2013 World Congress on Computer and Information Technology (WCCIT)*. IEEE, 2013, pp. 1–9.

[13] M. Kedadouche, M. Thomas, and A. Tahan, "Empirical mode decomposition of acoustic emission for early detection of bearing defects," in *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*. Springer, 2014, pp. 367–377.

[14] A. P. Engelbrecht, *Computational intelligence: an introduction*. Wiley-Blackwell, 2007.

[15] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2009.

[16] T. M. Mitchell, *Machine Learning*. McGraw Hill, 1997.

[17] A. M. Al Tobi and I. Duncan, "Improving intrusion detection model prediction by threshold adaptation," *Information*, vol. 10, no. 5, 2019.

[18] L. Raczkowski, M. Mozejko, J. Zambonelli, and E. Szczurek, "Ara: accurate, reliable and active histopathological image classification framework with bayesian deep learning," *Scientific Reports*, vol. 9, 2019.

[19] W. Khan, D. Ansell, K. Kuru, and M. Bilal, "Flight guardian: autonomous flight safety improvement by monitoring aircraft cockpit instruments," *Journal of Aerospace Information Systems*, vol. 15, no. 4, pp. 203–214, 2018.

[20] R. H. Torres, O. Ohashi, G. Garcia, F. Rocha, H. Azpúrua, and G. Pessin, "Exploiting machine learning models to avoid texting while driving," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.

[21] E. Fernandes, R. L. Rocha, B. Ferreira, E. Carvalho, A. C. Siravenha, A. C. S. Gomes, S. Carvalho, and C. R. B. de Souza, "An ensemble of convolutional neural networks for unbalanced datasets: A case study with wagon component inspection," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–6.

[22] G. Garcia, F. Rocha, M. Torre, W. Serrantola, F. Lizarralde, A. Franca, G. Pessin, and G. Freitas, "Rosi: A novel robotic method for belt conveyor structures inspection," in *2019 19th International Conference on Advanced Robotics (ICAR)*, 2019, pp. 326–331.

[23] P. Merriaux, R. Rossi, R. Boutteau, V. Vauchey, L. Qin, P. Chanuc, F. Rigaud, F. Roger, B. Decoux, and X. Savatier, "The vikings autonomous inspection robot: Competing in the argos challenge," *IEEE Robotics Automation Magazine*, vol. 26, no. 1, pp. 21–34, 2019.

[24] M. Tanaka, K. Kon, M. Nakajima, N. Matsumoto, S. Fukumura, K. Fukui, H. Sawabe, M. Fujita, and K. Tadakuma, "Development and field test of the articulated mobile robot t2 snake-4 for plant disaster prevention," *Advanced Robotics*, vol. 34, no. 2, pp. 70–88, 2020.

[25] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, pp. 93 – 101, 2019.

[26] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, pp. 197 – 227, 2016.

[27] F. T. Liu, K. M. Ting, and W. Fan, "Maximizing tree diversity by building complete-random decision trees," in *Advances in Knowledge Discovery and Data Mining*, T. B. Ho, D. Cheung, and H. Liu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 605–610.

[28] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, p. 16, 2018.

[29] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133–144, 2018.

[30] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, 2019.

[31] B. Safa, T. J. Arkebauer, Q. Zhu, A. Suyker, and S. Irmak, "Latent heat and sensible heat flux simulation in maize using artificial neural networks," *Computers and electronics in agriculture*, vol. 154, pp. 155–164, 2018.

[32] B. P. Lathi, *Signal processing and linear systems*. Oxford University Press New York, 1998.

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, 2009.