27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

# COB-2023-2267
# COMPARISON OF MACHINE LEARNING TECHNIQUES FOR FAULT DIAGNOSIS IN BELT CONVEYOR IDLERS

**João Lucas Lobato Soares**
**Thiago Barroso Costa**
**Lis Silva de Moura**
**Walter dos Santos Sousa**
**Alexandre Luiz Amarante Mesquita**
**André Luiz Amarante Mesquita**
Laboratory of Fluid Dynamics and Particulate, FluidPar, Federal University of Pará, Av. Brasília, s/n, Vila Permanente, 68.455-901, Tucuruí, PA
soaresjoao974@gmail.com
thiago.costa@tucurui.ufpa.br
lis.moura@tucurui.ufpa.br
wss@ufpa.br
alexmesq@ufpa.br
andream@ufpa.br

**Danilo de Souza Braga**
Dynamox, Rd. SC 401, 600, St. Parque Tecnológico Alfa, Florianópolis - SC, 88.030-909
danilo@dynamox.net

***Abstract.*** *Belt conveyors are essential components in the mining industry responsible for transporting ore. However, failures in their components can compromise the entire production process, leading to unexpected stops and delays in the schedule. Therefore, it is crucial to perform regular diagnostics and replace damaged parts to avoid these problems. Idler failure represents one of the primary failure modes in conveyor belt components, mainly associated with wear and bearing defects. Machine learning-based techniques have been successfully applied for condition monitoring and fault diagnosis in these components. Moreover, the objective of this study is to develop a fault diagnosis system for conveyor belt idlers utilizing artificial intelligence techniques, while also conducting a comparative analysis of the performance of proposed classification methods. The first method consists of applying the Gradient Boosting Decision Tree (GBDT), and the second consists of applying the Multilayer Perceptron (MLP) to the energy bands from the application of Wavelet Packet Decomposition (WPD) to the vibration signals of the belt conveyor. The performances of the methods are compared for each failure mode. The results obtained showed that for bearing defects GBDT presented accuracy above 100% from 30 estimators, while MLP obtained maximum accuracy of 98.3% with 100 iterations, which indicates better performance with decision trees application. Furthermore, for surface wear GBDT obtained maximum accuracy of 98.3% from 21 estimators, while MLP reached 96.7% with 100 iterations. Therefore, it is inferred that GBDT presented to be a more effective model compared to MLP by initial parameters. However, both models have high accuracy as classifiers for detection of idler failures.*

***Keywords:*** *machine learning, fault diagnosis, GBDT, MLP, belt conveyor idler*

## 1. INTRODUCTION

Belt conveyors are commonly used for bulk material transportation and require integrated maintenance systems across all components to enhance asset reliability. Among these components, idlers are one of the main focal points for study, evaluating operational conditions, reliability estimates, and monitoring data to improve decision-making accuracy. Consequently, predictive maintenance is an important tool for ensuring the integrity of the belt conveyor system through monitoring physical parameters such as vibration and temperature (Liu et al., 2019).

Vibration analysis is thus one of the primary fault detection tools employing signal monitoring techniques for rotating machinery. Signal patterns are established when the equipment is under normal conditions, and subsequent deviations in the time, frequency, or time-frequency domains are observed to diagnose potential component faults (Popescu et al., 2021). Furthermore, predictive maintenance has emerged as a major paradigm of Industry 4.0, facilitating increased communication among remote monitoring systems with enhanced processing and memory capabilities (Muniz et al., 2023).

However, even with remote monitoring, predictive inspection by workers becomes impractical due to the number of idlers installed along the extensive length of the belt conveyor. Therefore, machine learning emerges as an alternative for diagnosing faults through vibration analysis. Machine learning models create classifiers for signal characteristics and accurately determine the component's health (Alharbi et al., 2023). Gradient Boosting Decision Tree (GBDT) (Liu et al., 2020) and Artificial Neural Network (ANN) (Ravikumar et al., 2020) are among the primary machine learning techniques already applied in the literature for idlers.

Nevertheless, due to the non-linear and non-stationary nature of vibration signals, traditional features extracted from the time or frequency domain alone are not highly effective in diagnosing idlers. It becomes necessary to employ time-frequency domain techniques such as Wavelet Packet Decomposition (WPD). WPD is an extension of the Wavelet Transform that extracts signal energy from frequency bands, enabling the identification of disparities between healthy and faulty signals (Li et al., 2013).

Therefore, in this study, vibration signals from belt conveyor idlers were decomposed into wavelet energy bands, followed by a comparison of machine learning models. The chosen techniques for analysis were GBDT and ANN to identify which technique exhibits superior accuracy improvement.

## 2. WAVELET PACKET DECOMPOSITION

The wavelet transform can be analytically described as a set function, as represented in Eq. (1), which is simplified to a single function. This function involves a set of dilation and translation operations, forming families of wavelets. These families are chosen as mathematical tools in various fields (Daubechies, 1988). In the equation, n represents the level of decomposition, j is the scale factor, and k is the translation factor.

Alternatively, Multiresolution Analysis (MRA) is employed as a viable approach to wavelet analysis, enabling the creation of functions based on specific conditions tailored to the characteristics of the signals under investigation (Daubechies, 1988; Rowe and Abbott, 1995). The Daubechies wavelet family was developed based on MRA, which lacks an analytical expression and is obtained through algorithms designed to fulfill the required analysis properties (Torrence and Compo, 1998).

$$W_{j,k}^{n}(t) = 2^{\frac{j}{2}} W^{n}(2^{j}t - k) \tag{1}$$

The Wavelet Transform (WT) utilizes a wavelet basis, similar to the sine functions used in the Fourier Transform. While sine functions are periodic and have a constant amplitude throughout the domain, the wavelet basis consists of short, periodic functions with a linear value of zero outside their domain (Torrence and Compo, 1998). Wavelet Packet Decomposition (WPD) is an extension of WT, where the signal is filtered by a sequence of high-pass (HP) and low-pass (LP) filters, depending on the analyzed decomposition level. Fig. (1) illustrates the functioning process of WPD for two decomposition levels (Wang et al., 2013).

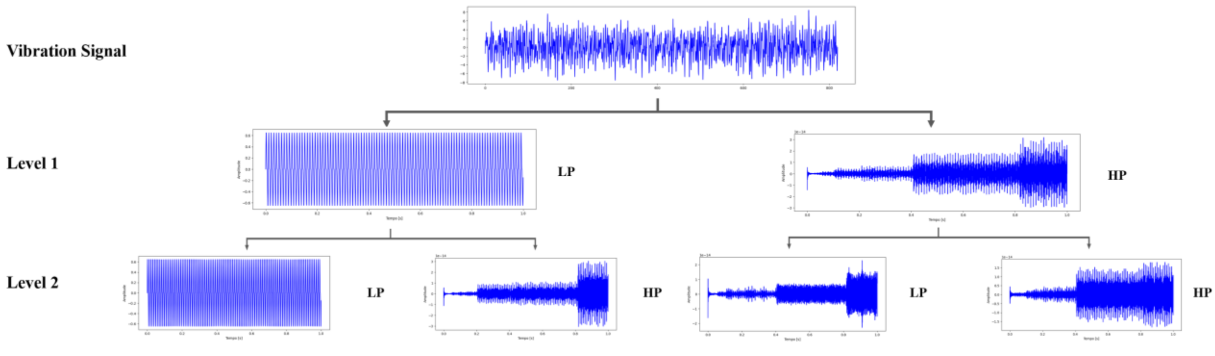


Figure 1. Vibration signal decomposition.

Therefore, WPD provides the coefficients of the frequency bands of a vibration signal, which can be used as features to identify faults through wavelet band feature extraction in an intelligent classification algorithm. Eqs. (2) and (3) below represent the Wavelet Packet coefficients and the band energy, respectively, where $f(t)$ represents the signal in the time domain (Li et al., 2013).

$$w_{j,k}^{n} = \int_{-\infty}^{+\infty} f(t)\, W_{j,k}^{n}\, dt \tag{2}$$

$$E(J,i) = \left\| w_{j,k}^{n} \right\|^{2} \tag{3}$$

In these equations, the Wavelet Packet coefficients are represented by $w_{j,k}^{n}$, and the band energy is calculated as the sum of the squares of these coefficients.

## 3. MACHINE LEARNING TECHNIQUES

Machine learning is a scientific field of study that focuses on algorithms and statistical models where a dataset is processed and used to learn and perform specific tasks automatically and efficiently. One of the key types of learning is supervised learning. In supervised learning, a function is executed based on labeled training data, requiring external assistance. The algorithm learns to identify patterns in the data's features and classify them according to the provided labels (Mahesh, 2018).

In the context of diagnosing and prognosing mechanical failures in the industry, machine learning techniques have been widely studied. Literature reviews indicate that 53.3% of published studies in this field apply supervised learning techniques (Fernandes et al., 2022). Furthermore, several techniques have been proposed for fault detection in belt conveyor idlers using vibration signals collected from the components. Below are some of these techniques.

### 3.1 Gradient Boosting Decision Tree

Decision trees are considered non-parametric methodologies, and their interpretation is based on partitions called nodes. Initially, each partition defines a condition on the distribution of the data by classes and divides the samples into other nodes, conditioning them successively according to the established learning parameters. The final nodes of the tree are called leaves (Izbicki and Santos, 2020).

One of the main advantages of decision trees is the ease of interpreting predictions, which can be visualized in the form of graphs. Additionally, trees can handle large-scale datasets and can be optimized using methods that combine multiple trees to improve prediction accuracy. One of the main optimization methods is Gradient Boosting Decision Trees (GBDT). With GBDT, it is possible to enhance the accuracy of decision trees through classification rules that minimize the learning error (Liu et al., 2020).

According to Liu et al. (2021), a weak learner is found, in which a decision tree model is fit to the residuals of a previous model to minimize the error between the output values and the true values. Thus, the final predictive model is obtained by summing the results of all models from the previous iteration. The GBDT process involves finding the optimized parameters α and β through Eq. (4), where $F_0(x_i)$ is the prediction function obtained at the m-th iteration. Initially, an initial weak classifier $F_0$ is defined (Eq. (5)), starting from the constant β and reaching the minimum value of the loss function $L(y, F_m(x))$ (Liu et al., 2020).

$$(\alpha_m, \beta_m) = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} (L(y_i, F_m(x_i)) + \beta h(x_i; \alpha)) \tag{4}$$

$$F_0 = \arg\min_{\beta} \sum_{i=1}^{N} (L(y_i, \beta)) \tag{5}$$

To reduce the loss function, the m-th classifier ($\beta_m h(x;\ \alpha_m)$) is constructed in the downhill direction of the gradient, as indicated in Eq. (6). The sample data is fitted using the base classifiers to obtain an initial model, along with the parameter $\alpha_m$ and the fit $h(x;\ \alpha)$ in Eq. (7). With $\alpha_m$ calculated, Eq. (8) is used to update the current model's weight. With the established parameters, the prediction function (Eq. (9)) is calculated and updated after each iteration until a convergence condition is achieved (Liu et al., 2020).

$$-g_m(x_i) = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right], F(x_i) = F_{m-1}(x_i), i = 1, \dots, N \tag{6}$$

$$\alpha_m = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} [-g_m(x_i) - \beta h(x_i; \alpha)]^2 \tag{7}$$

$$\beta_m = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} L([y_i, F_{m-1}(x_i) + \beta h(x_i; \alpha_m)]^2) \tag{8}$$

$$F_m(x) = F_m(x) + \beta_m h(x_i; \alpha_m) \tag{9}$$

### 3.2 Artificial Neural Network – Multilayer Perceptron

Artificial Neural Networks (ANN) are a learning technique based on modeling the neurons of the nervous system. Neurons in an ANN are organized into layers and communicate through activation functions. The model consists of three types of layers: the input layer (containing the initial variables of the dataset), the output layer (providing the final response of the neural network), and the hidden layers (intermediate layers that optimize communication between input and output). One of the most commonly used ANN techniques is the Multilayer Perceptron (MLP), which is applied to create supervised learning models using the Backpropagation algorithm (Ravikumar et al., 2020).

The Backpropagation algorithm involves an iterative process of reducing the error in the output layer. The neural network weights, applied in nonlinear activation functions, are updated through a gradient descent of the loss function to previous layers. Equation (10) presents the weight vector (w) update process in the k-th iteration. The symbol ∇E represents the gradient of the loss function with respect to the weights, and δ is the learning rate of the model (Brunton and Kutz, 2019)

$$w_{k+1} = w_k + \delta \nabla E \tag{10}$$

## 4. EXPERIMENTAL/COMPUTATIONAL PROCEDURE

The work was divided into three main stages, as illustrated in Fig. (2). Initially, the idlers were installed on the conveyor in four rows of supports, with each support containing a set of three rollers. These rollers were angularly positioned on the sides of the belt at a 45° angled, and one idler was positioned horizontally under the central region of the belt. After the installation and alignment of the idlers and belt, a vibration sensor was placed on the lateral region of the support, as shown in Fig. (3), to measure the signals after the conveyor's activation. The used sensor was a Teknikao® accelerometer, model NK 30, with an output of 104 mV/g.


Figure 2. Flow chart for fault diagnosis.

Figure 3. Experimental apparatus.

The accelerometer is connected to the Teknikao® signal analyzer, and the signals are processed by the manufacturer's own software called SDAV (Digital Vibration Analysis System). The initial program settings allow the display of velocity and acceleration vibration signals in two distinct windows. The test time for each window is approximately 6.55 seconds, resulting in a total of 32,000 captured data points. The vibration frequency range was configured from 1 Hz to 2 kHz.

For experimentation purposes, two types of rollers with different failure modes were used, as illustrated in Fig. (4). In one roller, the bearings located at the ends of the inner cylinder were damaged, while in the other roller, the shell surface that comes into direct contact with the belt exhibited wear. For each test, one defective roller was installed in one of the lateral regions of the support, while the remaining rollers maintained normal conditions. In total, 240 tests were conducted, with 80 performed under normal conditions, 80 with a idler having defective bearings, and 80 with a idler showing surface wear.

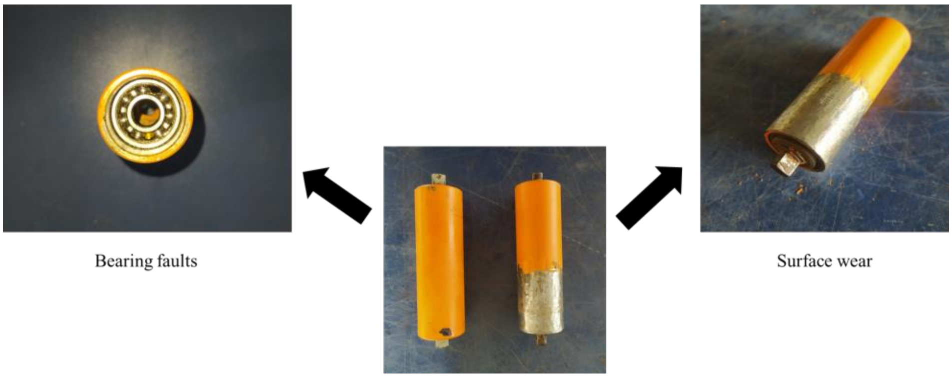


Figure 4. Failure mode.

After collecting the signals obtained from the tests, each signal was divided into 4 samples, resulting in a database with 960 vibration signal samples, each approximately 1.64 seconds long. The database was processed using a Machine Learning algorithm developed in Python 3.7®. For vibration signal analysis, the signals were decomposed into 15 levels of decomposition using the 'db 8' wavelet function, which is part of the Daubechies wavelet family. The technique used was Wavelet Packet Decomposition (WPD), which allowed the calculation of wavelet coefficients and energy for each resulting band from the decomposition. Furthermore, the wavelet energy bands were preprocessed and normalized to fit new values within a normalized range.

From the normalized wavelet energy, a new database was created in which each energy band represented a feature of the sample. Additionally, for each sample, the signal state was identified as either 'normal condition' or 'faulty condition', represented respectively by '0' and '1' in the database. The dataset was then used to train fault diagnostic models using the GBDT (Gradient Boosting Decision Trees) and ANN-MLP (Artificial Neural Network - Multilayer Perceptron) techniques. The initial parameters used are presented in Table (1). The data was divided into training and testing sets, with a ratio of 75% for training and 25% for testing. Additionally, cross-validation technique was applied to the training data, subdividing it into five groups.

Table 1. Machine learning techniques parameters

| **Parameter** | **Status** |
|---|---|
| **GBDT** | |
| Learning rate | 1.0 |
| Maximum depth | 2 |
| Random state | None |
| Number of iterations | 1 - 50 |
| **MLP** | |
| Number of hidden layers | 2 |
| Number of neuron/hidden layer | 10 |
| Activatio function | ReLU / Sigmoide |
| Optimizer | Adam |
| Loss function | Cross entropy |
| Number of iterations | 1 - 100 |

## 5. RESULTS

With the model created by GBDT, the accuracy of the models was extracted in comparison with the variation of the number of estimators for both defects. As observed in Figs. (5) and (6), there was higher convergence of diagnostic data for bearing defects compared to wear diagnosis. In all cases, models were created that achieved an accuracy above 98% out of the 50 applied tests. However, only the tests for bearing defects reached 100% accuracy with mp = 2. As seen in Fig. (5), from four estimators, the model achieved accuracy above 95% for the test data, and variations between 99.5% and 100% from 30 estimators, indicating that the classification was well adjusted between training and testing. Additionally, the training data reached 100% accuracy from 16 estimators, which indicates the minimum value to be adopted as a hyperparameter.

Regarding the model created for the diagnosis of idler surface wear, there was no convergence of 100% accuracy for the test data, considering the same hyperparameter setup. However, it started to perform above 95% accuracy from 10 estimators and with a maximum accuracy of approximately 98.3% with 21 GBDT boosting stages.

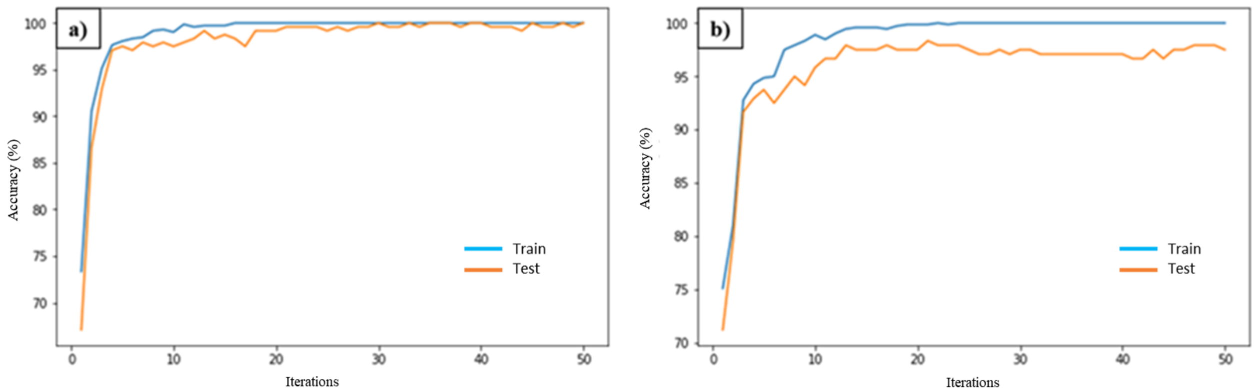


Figure 5. Accuracy x iterations (GBDT). a) Bearing faults. b) Surface wear.

With the establishment of the model's performance in relation to the number of estimators for both failure modes, confusion matrices were created for bearing defects and wear, represented by Tab. (2) and (3), respectively. As 100% accuracy was not achieved for the diagnosis of surface wear, the secondary diagonal of the matrix represents the quantity of incorrect predictions generated by the model, with five errors for faulty conditions and one error for normal condition. The minimization of prediction errors is supported by the fact that the omission in diagnosing a roller defect (false negative) occurred only once, which, in practice, may indicate a mitigated problem. It should be noted that situations of false positives can be identified through manual inspections of the equipment.

Table 2. Confusion matrix (GBDT – Bearing faults)

| | Normal condition | Faulty condition | |
|---|---|---|---|
| Normal condition | 128 | 0 | Real |
| Faulty condition | 0 | 112 | |
| | Prediction | | |

Table 3. Confusion matrix (GBDT – Surface wear)

| | Normal condition | Faulty condition | |
|---|---|---|---|
| Normal condition | 115 | 5 | Real |
| Faulty condition | 1 | 119 | |
| | Prediction | | |

With the results generated by the machine learning pattern with gradient, the tree created with the rules established by GBDT was analyzed, considering the model's loss function optimization. As perceptible in Fig. (6), for both failure modes, nodes were created with rules related to the 13th wavelet energy band.

The difference is that for bearing defects, this energy band is present in the root node rule, while for wear defects, parent nodes with the 13th band were created for leaf branching. Additionally, the tenth band was applied in the model as a parent node for bearing defect detection. For idler wear detection, the 12th band was determined as the rule in the root node of the decision tree.

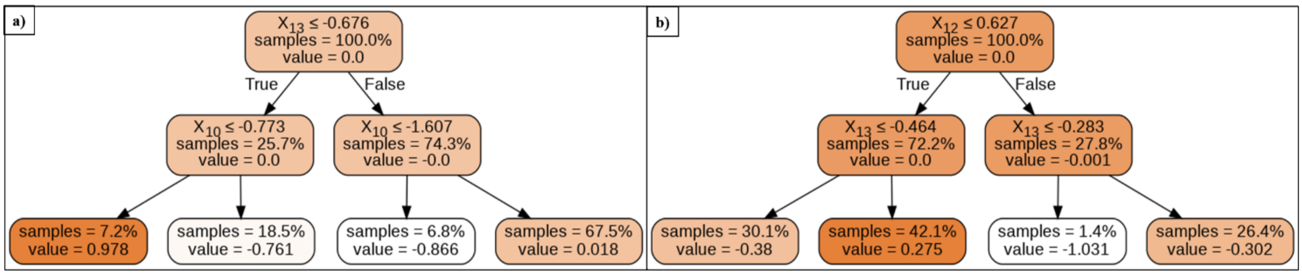


Figure 6. Decision trees (GBDT). a) Bearing faults. b) Surface wear.

Just like in the GBDT analysis, in MLP, the results presented better models for bearing fault detection compared to wear detection. For each failure mode, the algorithm performed 100 iterations of Backpropagation, represented by Figs. (7) and (8). However, there is less overfitting observed for wear defects, indicating a more accurate model despite lower exactude.

In Fig. (7), it was noticed that until the ninth iteration, the test results achieved better accuracies compared to the training. Furthermore, training started to show accuracy above 90% from 38 iterations, while for testing, accuracy reached a value above 90% from 49 iterations. The best bearing fault diagnostic models achieved accuracies of approximately 99.4% and 98.3% for training and testing data, respectively. This indicates that the best model was obtained with 100 iterations.

It can also be observed in Fig. (8) that between 1 and 65 iterations, accuracy values performed better on the test set compared to training. As a result, test accuracy reached 90% with 47 iterations, while training only reached 90% from 54 iterations. The best models had performances of approximately 97.0% and 96.7% accuracy for training and testing, respectively. For wear detection, the best learning model was also obtained with 100 iterations.

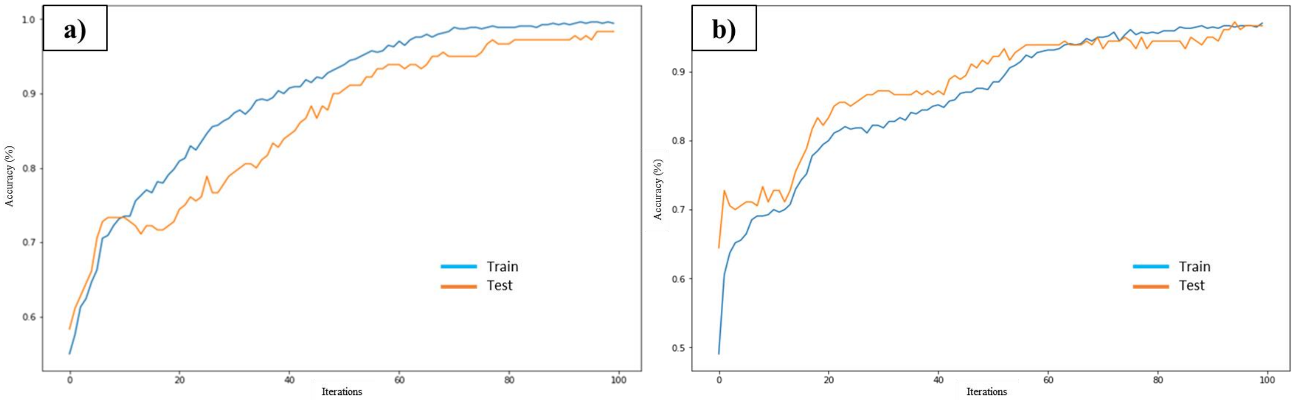


Figure 7. Accuracy x iterations (MLP). a) Bearing faults. b) Surface wear.

After analyzing the best model, confusion matrices for both failure modes were plotted in Tabs. (4) and (5). For bearing defects, it is important to note that there were more errors for false positives than false negatives, reducing the number of undiagnosed failures in the belt conveyor and minimizing unplanned downtime. This false negative error represents only 0.89% of the model's prediction failure. For wear defects, there was an equal number of errors in false positive and false negative diagnoses on the secondary diagonal, with a prediction failure of 3.4% and 3.3% for both errors, respectively.

Table 4. Confusion matrix (MLP – Bearing faults)

| | Normal condition | Faulty condition | |
|---|---|---|---|
| Normal condition | 111 | 6 | Real |
| Faulty condition | 1 | 122 | |
| | Prediction | | |

Table 5. Confusion matrix (MLP – Surface wear)

| | Normal condition | Faulty condition | |
|---|---|---|---|
| Normal condition | 117 | 4 | Real |
| Faulty condition | 4 | 115 | |
| | Prediction | | |

## 6. CONCLUSIONS

In conclusion, the implementation of fault diagnosis models for idlers using GBDT and MLP classification techniques yielded satisfactory results. The learning models achieved accuracy above 98% for both defects when utilizing GBDT. Moreover, the Multilayer Perceptron demonstrated a good fit between training and testing, indicating the creation of a satisfactory learning model for both failure modes. Remarkably, the MLP technique consistently achieved accuracies above 96% across all cases for 100 iterations. As an opportunity for future work, laboratory tests of the loaded belt conveyor should be conducted to further validate the models' effectiveness under realistic operating conditions. Additionally, the treatment and analysis of real data from conveyors supervised by Dynamox® should be initiated. This will allow for the refinement and enhancement of the fault diagnosis techniques, ultimately improving the maintenance strategies and preventing unforeseen failures in conveyor systems supervised by Dynamox®.

## 7. ACKNOWLEDGEMENTS

This paper was supported by Dynamox S.A, where their expertise and assistance have been instrumental in enhancing the quality and clarity of our research, which have significantly contributed to the success of this project.

## 8. REFERENCES

Alharbi, F., Luo, S., Zhang, H., Shaukat, K., Yang, G., Wheeler, C. A., & Chen, Z. (2023). A brief review of acoustic and vibration signal-based fault detection for belt conveyor idlers using machine learning models. Sensors, 23(4), 1902.

Brunton, S. L., & Kutz, J. N. (2022). Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. Communications on pure and applied mathematics, 41(7), 909-996.

Fernandes, M., Corchado, J. M., & Marreiros, G. (2022). Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Applied Intelligence, 52(12), 14246-14280.

Izbicki, R., & dos Santos, T. M. (2020). Aprendizado de máquina: uma abordagem estatística. Rafael Izbicki.

Li, W., Wang, Z., Zhu, Z., Zhou, G., & Chen, G. (2013). Design of online monitoring and fault diagnosis system for belt conveyors based on wavelet packet decomposition and support vector machine. Advances in Mechanical Engineering, 5, 797183.

Liu, X., He, D., Lodewijks, G., Pang, Y., & Mei, J. (2019). Integrated decision making for predictive maintenance of belt conveyor systems. Reliability Engineering & System Safety, 188, 347-351.

Liu, X., Pei, D., Lodewijks, G., Zhao, Z., & Mei, J. (2020). Acoustic signal based fault detection on belt conveyor idlers using machine learning. Advanced Powder Technology, 31(7), 2689-2698.

Liu, Y., Miao, C., Li, X., Ji, J., & Meng, D. (2021). Research on the fault analysis method of belt conveyor idlers based on sound and thermal infrared image features. Measurement, 186, 110177.

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

Muñiz, R., Nuño, F., Díaz, J., González, M., J Prieto, M., & Menéndez, Ó. (2023). Real-time monitoring solution with vibration analysis for industry 4.0 ventilation systems. The Journal of Supercomputing, 79(6), 6203-6227.

Popescu, T. D., Aiordachioaie, D., & Culea-Florescu, A. (2021). Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview. The International Journal of Advanced Manufacturing Technology, 1-17. Popescu, T. D., Aiordachioaie, D., & Culea-Florescu, A. (2021). Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview. The International Journal of Advanced Manufacturing Technology, 1-17.

Ravikumar, S., Kanagasabapathy, H., & Muralidharan, V. (2020). Multi-component fault diagnosis of Self Aligning Troughing Roller (SATR) in belt conveyor system using decision tree: A statistical approach. FME Transactions, 48(2), 364-371.

Rowe, A. C., & Abbott, P. C. (1995). Daubechies wavelets and mathematica. Computers in Physics, 9(6), 635-648.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological society, 79(1), 61-78.

Wang, X., Liu, C., Bi, F., Bi, X., & Shao, K. (2013). Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension. Mechanical Systems and Signal Processing, 41(1-2), 581-597.

## 9. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.