

COB-2023-2261

DECISION TREE FOR FEATURE SELECTION TO DIAGNOSIS OF BELT CONVEYOR IDLER

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, as an essential equipment in the mining industry, requires constant monitoring to maintain good reliability. In order to support the belt and the transported material, idlers are components that, constantly, fail during operation, in which they present bearing defects as the most common failure modes. These defects can propagate more severe failures on the conveyor such as wear or tear on the belt, if idler becomes stuck. Thus, monitoring based on predictive maintenance is essential, and machine learning techniques can be used as an alternative for detecting equipment failures. In diagnostics using machine learning, the feature selection step is important to avoid overfitting and loss of accuracy in the classification of the equipment's condition. The present study analyzes the performance of decision tree algorithm as alternative method for dimensionality reduction. Initially, vibration signals were collected on the belt conveyor bench idler and the Wavelet Packet Decomposition (WPD) was applied to the signals to obtain the energy bands, which were used as features for classification. After the features were determined, three methods were analyzed for feature selection: application of no dimensionality reduction method, application of traditional Principal Component Analysis (PCA), and application of decision tree. Additionally, different classification algorithms were employed: Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN). As results, it was found a superior performance of diagnosis accuracy in all techniques with a dimensionality reduction of features selected by decision tree. Moreover, SVM, KNN and ANN presented accuracy above 98%, indicating high efficiency of the proposed decision tree application as a feature dimensionality reduction for classification.

Keywords: machine learning, fault diagnosis, feature selection, decision tree, belt conveyor idler

1. INTRODUCTION

Belt conveyors are widely used equipment for bulk material transportation and require integrated maintenance systems for all components in order to increase asset reliability. Among these components, the idler is one of the main targets of study, as it involves the evaluation of operating conditions, reliability estimates, and monitoring data to improve decision-making accuracy. Therefore, predictive maintenance is an important tool to ensure the integrity of the conveyor system through the monitoring of physical parameters such as vibration and temperature (Liu et al., 2019).

Vibration analysis is one of the primary fault detection tools using techniques for monitoring signals from rotating machinery. Signal patterns are established when the equipment is under normal conditions, and subsequently, changes in the time domain, frequency, or time-frequency domain are observed to diagnose possible component failures (Popescu et al., 2021). Furthermore, predictive maintenance has become one of the key paradigms of Industry 4.0, enabling greater communication between remote monitoring systems with enhanced processing and memory capabilities (Muniz et al., 2023).

Moreover, even with remote monitoring, predictive inspection by workers becomes impractical due to the large number of idlers installed along the extensive length of the conveyor belt. Therefore, machine learning emerges as an alternative for diagnosing failures through vibration analysis. Machine learning creates models that classify signal characteristics and determine component health with high accuracy (Alharbi et al., 2023). Among the main machine learning techniques already applied in the literature for idlers, notable examples include Support Vector Machine (SVM) (Li et al., 2013), Linear Discriminant Analysis (LDA) (Liu et al., 2021), K-Nearest Neighbors (KNN) (Yang et al., 2020), and Artificial Neural Network (ANN) (Ravikumar et al., 2020).

Furthermore, for model application, it is necessary to select the most significant features to reduce information redundancy and improve classification effectiveness. Feature selection aids in reducing the dimensionality of the database through various techniques that enhance the discriminatory power of the model while preserving a subset of features with greater relevance (Rauber et al., 2014). Ravikumar et al. (2020) applied decision trees to identify statistical features with greater capacity to diagnose faults in conveyor belt idlers. Decision trees highlight rules that divide the sample set into classes based on certain conditions of a feature.

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

Therefore, in this paper, vibration signals from conveyor belt idlers were decomposed into wavelet energy bands, from which the most relevant energy bands were identified as feature selection using decision trees. Subsequently, a comparison was made between machine learning models with and without feature selection. The selected techniques for analysis were SVM, LDA, KNN, and ANN, to identify which technique shows greater improvement in accuracy with the decision tree.

2. FEATURE EXTRACTION AND SELECTION

Feature extraction and selection allow input variables to be processed and adapted so that the learning model can classify the data with greater accuracy. Therefore, the main concepts of WPD and decision trees will be presented for feature extraction and selection, respectively.

2.1 Wavelet Packet Decomposition

Wavelet can be described analytically as a set of functions, as represented below in Eq. (1), simplified to a single function that undergoes a series of dilation and translation operations, forming wavelet families. These families are selected as mathematical tools in various fields (Daubechies, 1988). Here, n represents the decomposition level, j is the scale factor, and k is the translation factor.

On the other hand, Multiresolution Analysis (MRA) has been used as an alternative to wavelet analysis, allowing the creation of functions based on specific chosen conditions according to the characteristics of the signals being analyzed (Daubechies, 1988; Rowe and Abbott, 1995). The Daubechies function family was created based on MRA, which does not have an analytical expression and is obtained through algorithms designed to meet the required analysis properties (Torrence and Compo, 1998).

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

The Wavelet Transform (WT) applies the wavelet basis, analogous to the sinusoidal functions used in the Fourier Transform. While sine functions are periodic with constant amplitude beyond the domain, the wavelet basis is a short-periodic function with zero linear value outside the domain (Torrence and Compo, 1998). Wavelet Packet Decomposition (WPD) is an extension of the WT, where the signal is filtered by successive high-pass (HP) and low-pass (LP) filters according to the decomposition level of the analysis. Figure 1 represents the operation of WPD for 2 levels of decomposition (Wang et al., 2013).

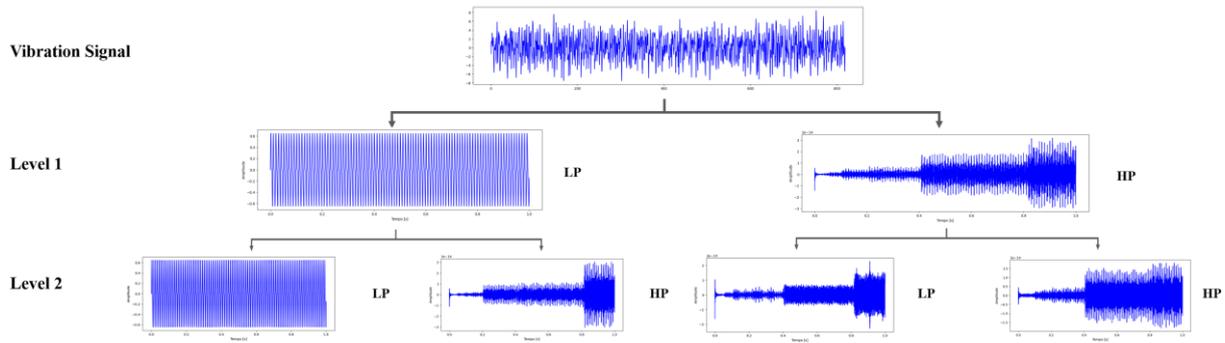


Figure 1. Vibration signal decomposition.

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

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

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

2.2 Decision Tree

For feature selection, decision trees can be applied to identify the most relevant features based on the level of individual contribution determined by the decision tree rules. Rules are established in nodes that classify samples into branches with leaves dividing the classes (Saravanan et al., 2009). The basic algorithm of decision trees involves surveying the features of the data vector x_j , as shown in Eq. (4), to identify the best prediction for the labels y_j (Eq. (5)). Then, a comparison of the accuracy of each feature is made, and the feature with the best prediction is selected for tree branching. Finally, further branching is performed with the new subgroups in each branch until the final grouping, called a leaf (Brunton and Kutz, 2019).

$$\{x_j \in \mathbb{R}^n, j \in \mathbb{Z} := \{1, 2, \dots, m\}\} \quad (4)$$

$$\{y_j \in \{\pm 1\}, j \in \mathbb{Z}' \subset \mathbb{Z}\} \quad (5)$$

By visualizing the tree, features that do not contribute to the model branching can be removed to enhance the classification effectiveness of machine learning.

3. MACHINE LEARNING TECHNIQUES

Machine learning is the scientific study of algorithms and statistical models in which a dataset is processed and learned to perform a specific task automatically and more efficiently. Among the main types of learning, supervised learning stands out. Supervised learning performs a specific function based on labeled training data, requiring external assistance. For each label, the algorithm learns to identify patterns in the data characteristics and classify them (Mahesh, 2018).

Perspectively, machine learning techniques have been studied for applications in the diagnosis and prognosis of mechanical failures in the industrial sector, and literature reviews indicate that 53.3% of published works have applied supervised learning (Fernandes et al., 2022). Additionally, techniques for detecting faults in conveyor belt idlers evaluated through vibration signals collected from the components have also been presented.

3.1 Support Vector Machine

The basic principle of SVM is to create an optimal data separation plane with linearly separable conditions between two classes, where the margin between the closest data points and the class boundary is known as support vectors (Yang et al., 2020). However, SVM can also solve non-linear and high-dimensional problems. Given a set of samples x_i and y_i

(for $x_i \in R^d$; $y_i \in \{-1, +1\}$; $i = 1, 2, \dots, n$), an algorithm is obtained through Lagrange's theorem, expressed by Eq. (6), where ω is the weight vector, b is the bias, ξ is the damping factor, and C is the penalty factor. Additionally, the Lagrange coefficients α_i satisfy the conditions shown in Eq. (7), where l is the number of support vectors (Li et al., 2013).

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i, \text{ s. t. } y_i[(\omega \cdot x_i) + b] \geq 1 - \xi_i \quad (6)$$

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \right\} \quad (7)$$

For $\alpha_i \neq 0$, the samples that satisfy the above conditions are the support vectors. Therefore, the classification function is obtained through Eq. (8). When the samples are not linearly separable, they can be mapped nonlinearly using the Kernel function $K(x_i, x_j)$. The main Kernel functions are linear, polynomial, RBF, and sigmoid (Li et al., 2013).

$$\alpha_i \{y_i[(\omega \cdot x_i) + b] - 1 + \xi_i\} = 0 \quad (8)$$

3.2 Linear Discriminant Analysis

LDA is a learning and dimensionality reduction technique that creates linear combinations from independent features extracted from data samples. These combinations calculate vectors in the underlying space that discriminate the classes more effectively (Martínez and Kak, 2001). Consequently, the discrimination aims to maximize the determinant ratio of the interclass and intraclass scatter matrices. The main difference between LDA and Principal Components Analysis (PCA) is that LDA does not generate pseudo features (Liu et al., 2021).

Considering the j -th sample and the i -th class forming an input vector x_j^i applied to the feature extraction vector f_j^i , the centroid is calculated as expressed in Eq. (9), where M_i is the number of samples in the i -th class. Additionally, the centroid for all features is calculated as expressed in Eq. (10), for c classes.

$$\mu_i = \frac{1}{M_i} \sum_{j=1}^{M_i} f_j^i \quad (9)$$

$$\mu = \frac{1}{M_i} \sum_{i=1}^c \sum_{j=1}^{M_i} f_j^i \quad (10)$$

Thus, the intra-class and inter-class scatter matrices are calculated, represented by S_w and S_b in Eqs. (11) and (12), respectively.

$$S_w = \sum_{i=1}^c \sum_{j=1}^{M_i} (f_j^i - \mu_i)(f_j^i - \mu_i)^T \quad (11)$$

$$S_b = \sum_{i=1}^c M_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (12)$$

Finally, the optimal LDA projection matrix is given by Eq. (13). Therefore, the objective function can be expressed by the feature applicability index in Eq. (14).

$$W_{opt} = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1, w_2, \dots, w_m] \quad (13)$$

$$L^{DA} = 10 \log_{10} \frac{|W^T S_b W|}{|W^T S_w W|} \quad (14)$$

3.3 K-Nearest Neighbors

KNN is one of the main unsupervised learning techniques based on the clustering method, where distances between the sample data and the created cluster centers are calculated. Initially, initial values of K cluster centers μ_1, \dots, μ_K , are created, which are labeled by C_j determined in Eq. (15), for the input data x_i of the i sample (Yang et al., 2020).

$$C_j = \arg \min_j \|x^{(i)} - \mu_j\|^2 \quad (15)$$

Thus, the cluster centers μ_j are updated according to Eq. (16), for a database of m samples. After calculating the new means, the new clusters reclassify the samples. It is important to note the importance of choosing the initial values of the cluster centers in order to optimize the machine learning process (Yang et al., 2020).

$$\mu_j = \frac{\sum_{i=1}^m 1\{c^{(i)} = j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}} \quad (16)$$

3.4 Artificial Neural Network

ANN is a learning technique based on modeling the nervous system neurons. The neurons in ANNs are arranged in layers and communicate through activation functions. The model is divided into three layer levels: the input layer (consisting of initial variables from the database), the output layer (providing a final response from neural communication), and the hidden layers (intermediate layers that optimize the communication between input and output). One of the main ANN techniques is Multilayer Perceptron (MLP). MLP is applied to create supervised learning models with Backpropagation algorithms. Backpropagation involves an iterative process of reducing the error of the output layer, where network weights applied in non-linear activation functions are updated based on a gradient descent of the loss function to the previous layers (Ravikumar et al., 2020).

Eq. (17) presents the process at the k -th iteration of updating the vector w denoting the weights of the activation functions. ∇E is the gradient of the loss function with respect to the weight function, and δ is the learning rate of the model (Brunton and Kutz, 2019).

$$w_{k+1} = w_k + \delta \nabla E \quad (17)$$

4. EXPERIMENTAL/COMPUTATIONAL PROCEDURE

The paper was developed in four main stages, as illustrated in Fig. (2). The idlers were installed on the conveyor in a sequence of four sets of three rollers each. Two rollers were positioned on the sides under 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 idler, as shown in Fig. (3), for signal measurement after the conveyor activation. The sensor used 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 was connected to the Teknikao® signal analyzer, and the signals were processed using the SDAV (Digital Vibration Analysis System) software developed by the manufacturer. The initial program settings processed the signal in two windows created to display velocity and acceleration vibration signals with a test time of approximately 6.55 seconds for each window, resulting in a total of 32,000 data points. In addition, the vibration frequency range was set between 1 Hz and 2 kHz.

A defect was introduced in one idler, as illustrated in Fig. (4), where the bearings located at the ends of the inner shell of the idler were damaged. For each test, the defective idler was installed in one of the lateral regions of the idler, while the other idlers remained in normal condition. A total of 160 tests were performed, with 80 samples having normal idlers and 80 samples with one idler having bearing defects.



Figure 4. Failure mode.

The signals obtained from the tests were collected, and each signal was divided into 4 samples, resulting in a database with 480 vibration signal samples, each approximately 1.64 seconds long. The database was acquired and processed by creating a Machine Learning algorithm using Python 3.7® language.

From the creation of a database, the vibration signals were decomposed into wavelet functions at 15 levels of decomposition for each sample. The selected wavelet function was the 'db 8' member extracted from the Daubechies family. WPD was applied to calculate the coefficients and wavelet energy for each band. Furthermore, the wavelet energy bands were preprocessed to adjust the samples to new values within a normalized range.

With the normalized wavelet energy calculated, a new database was extracted with each energy band representing a feature of the sample. Additionally, for each sample, the signal state was identified as 'normal condition' and 'defective condition', represented in the database as '0' and '1' for the two states, respectively.

After properly formatting the new database with normalized wavelet energy, 75% of the data samples were randomly used to select features through the decision tree. The objective was for the tree to indicate which features were applied to the nodes, while the unused features would not be applied in the learning model. The created model does not have a maximum depth, which implies that the tree will branch until the leaves group the samples with the lowest possible impurity. Once the selected features were defined for a new database, the data was split into training and testing sets. Therefore, the samples were divided in a 75%/25% proportion for training and testing data.

Finally, the data was trained using four distinct learning models for fault diagnosis. The applied models were SVM, LDA, KNN, and ANN. The parameters for each model are highlighted in Table (1). Based on the selected model, a comparison graph of accuracies was created to understand the effectiveness of feature selection using the decision tree.

Table 1. Machine learning techniques parameters

Model	Parameters		
SVM	C 1.0	Kernel RBF	Degree 3
LDA	Solver SVD	Number of components None [min (n_classes - 1, n_features)]	
KNN	Number of neighbors 5	Weights Uniform	
ANN – MLP	Solver LBFGS	Alpha 0,00001	Hidden layers sizes 15

5. RESULTS

Decision tree, represented in Fig. (5), shows the model formed using the dataset with 16 wavelet energies $x[k]$ for $k = \{0, 1, \dots, 15\}$ for each sample.

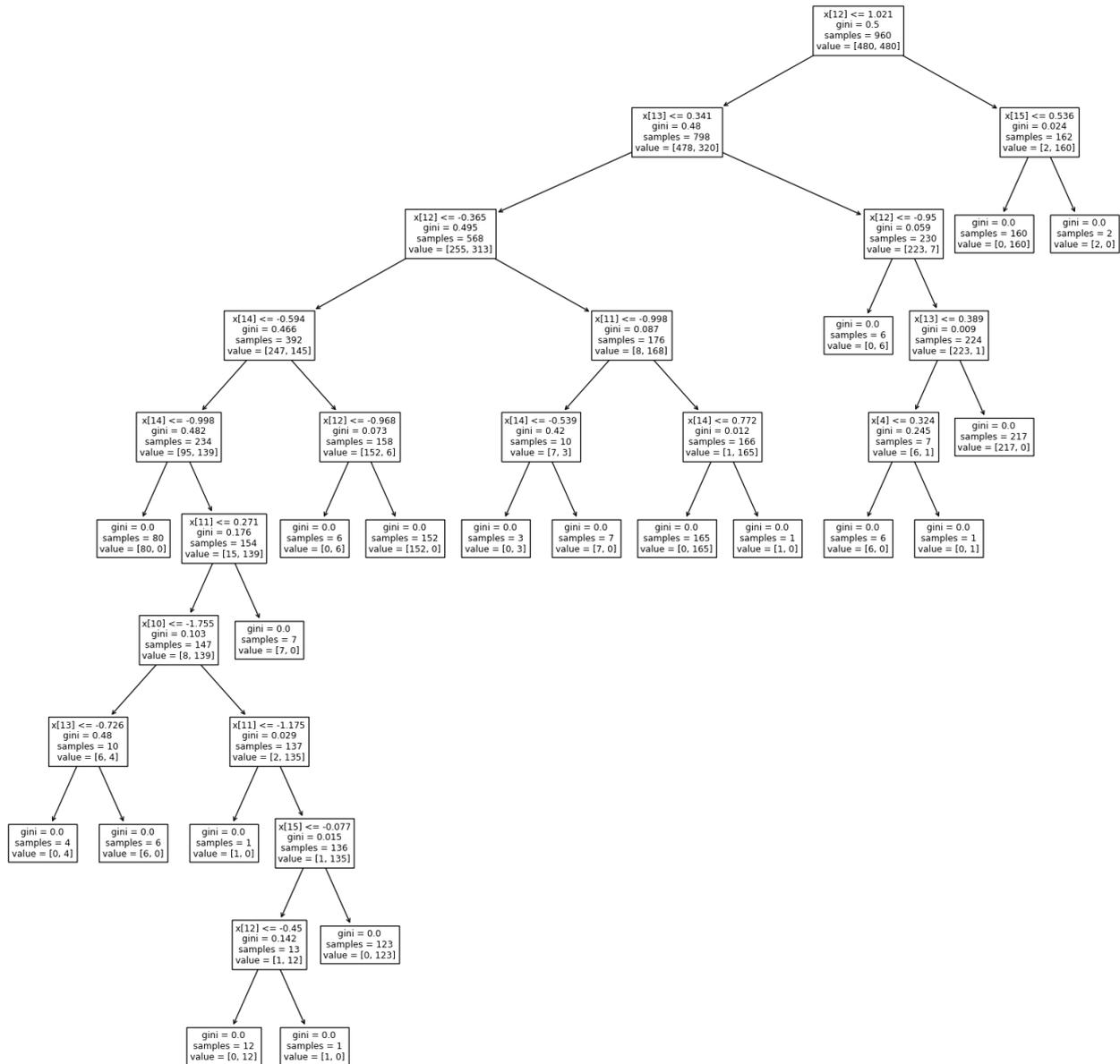


Figure 5. Decision tree.

As can be observed, the most relevant features in the tree were only 6 out of the 16 energies extracted from the samples, corresponding to energies from $x[10]$ to $x[15]$. It is evident that as the level of decomposition of the vibration signal increases using WPD, the identification of defective signals in the idlers becomes more noticeable. Therefore, models were created using only the 6 energies and compared to other models using all available energies. Fig. (6) presents the comparison of techniques with and without feature selection.

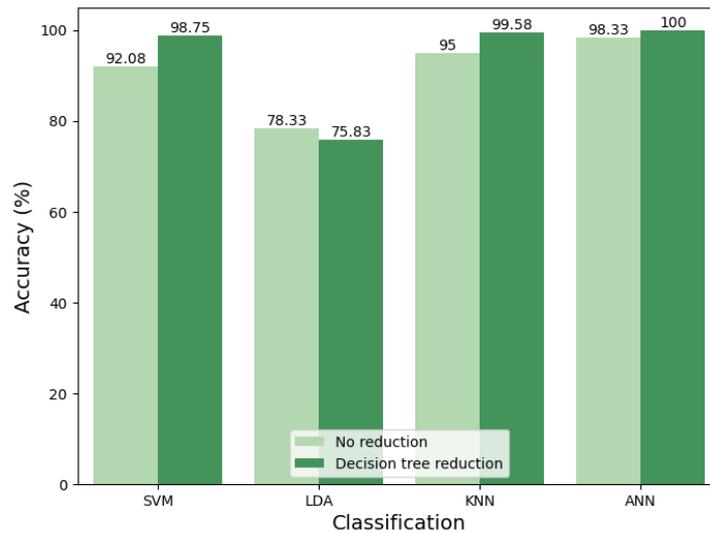


Figure 6. Comparative between learned techniques.

Among the applied techniques, the only one that did not show an improvement in classification performance was LDA, with a decrease in accuracies below 80%. This behavior is justified by the nonlinear characteristic of the vibration signals from the idlers, which makes it challenging to learn from the linear combinations of LDA. The other techniques showed a significant improvement with the implementation of decision trees for feature identification, with all of them achieving accuracies above 98% with feature selection. Among the three techniques, MLP-ANN stood out by reaching 100% accuracy after the application of the decision tree. However, there was only a 1.7% increase in accuracy. KNN achieved an accuracy of approximately 99.6% with a 4.8% increase, and SVM achieved an accuracy of 98.75% with a 7.2% increase, being the technique with the highest improvement.

6. CONCLUSIONS

With feature extraction and selection using WPD and the decision tree, respectively, there is a noticeable improvement in fault diagnosis accuracy using different machine learning techniques, especially nonlinear learning techniques such as SVM, KNN, and MLP-ANN, with correct classifications in over 98% of the tests. This behavior is justified by the nonlinear characteristic of the idler's vibration signal. Therefore, this method supports the effectiveness of applying decision trees for the most significant features.

For future opportunities, experimental vibration data should be collected from belt conveyor systems with materials to evaluate the influence of load on idlers under normal and defective conditions. Additionally, measurements should be taken from idlers with low-severity bearing defects to assess the detection effectiveness, particularly for incipient faults.

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.
- 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, 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.
- Rauber, T. W., de Assis Boldt, F., & Varejao, F. M. (2014). Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 62(1), 637-646.
- 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.
- Saravanan, N., Cholairajan, S., & Ramachandran, K. I. (2009). Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique. *Expert systems with applications*, 36(2), 3119-3135.
- 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.
- Yang, M., Zhou, W., & Song, T. (2020). Audio-based fault diagnosis for belt conveyor rollers. *Neurocomputing*, 397, 447-456.

9. RESPONSIBILITY NOTICE

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