signals

MDPI

Article

Decision Tree and ANOVA as Feature Selection from Vibration Signals to Improve the Diagnosis of Belt Conveyor Idlers

João L. L. Soares [1], Thiago B. Costa [1], Geovane S. do Nascimento [1], Walter S. Sousa [1], Jullyane M. S. de Figueiredo [2], Danilo S. Braga [2], André L. A. Mesquita [1] and Alexandre L. A. Mesquita [1,*]

[1] Laboratory of Fluid Dynamics and Particulate, FluidPar, Federal University of Pará, Tucuruí 68.455-901, PA, Brazil; soaresjoao974@gmail.com (J.L.L.S.); thiago.costa@tucurui.ufpa.br (T.B.C.); geovane.nascimento@tucurui.ufpa.br (G.S.d.N.); wss@ufpa.br (W.S.S.); andream@ufpa.br (A.L.A.M.)
[2] Product Department, Dynamox, Florianópolis 88.030-909, SC, Brazil; jullyane.figueiredo@dynamox.net (J.M.S.d.F.); danilo@dynamox.net (D.S.B.)
* Correspondence: alexmesq@ufpa.br

Abstract

This study aims to compare decision tree and Analysis of Variance (ANOVA) techniques as feature selection methods, combined with Wavelet Packet Decomposition (WPD) for feature extraction, to enhance the diagnosis of faults in belt conveyor idlers. Belt conveyors are widely used in mining for efficient transport, but idlers composed of rollers are frequently subject to failure, making continuous monitoring essential to ensure reliability. Automated diagnostic solutions using vibration signals and machine learning rely on signal processing for feature extraction, often requiring dimensionality reduction or feature selection to improve classification accuracy. Due to the limitations of traditional techniques such as Principal Component Analysis (PCA) in handling temporal variations, Decision Tree and ANOVA emerge as effective alternatives for feature selection. This framework applied to each feature selection method, and Support Vector Machine (SVM) was used as a classification technique. The diagnostic performance of each method, including the case without feature selection, was evaluated. The results showed a higher diagnostic accuracy performance for the approaches that applied the features from the decision tree and from ANOVA. The improvement in the diagnosis of roller failures with feature selection was corroborated with the hit rates of failure mode, severity level, and location of a defective roller above 93.5%.

Keywords: belt conveyor idlers; decision tree; ANOVA; fault diagnosis; feature selection

check for updates

Academic Editor: Manuel Duarte Ortigueira

Received: 17 June 2025
Revised: 31 July 2025
Accepted: 5 August 2025
Published: 13 August 2025

Citation: Soares, J.L.L.; Costa, T.B.; do Nascimento, G.S.; Sousa, W.S.; de Figueiredo, J.M.S.; Braga, D.S.; Mesquita, A.L.A.; Mesquita, A.L.A. Decision Tree and ANOVA as Feature Selection from Vibration Signals to Improve the Diagnosis of Belt Conveyor Idlers. *Signals* **2025**, *6*, 42. https://doi.org/10.3390/signals6030042

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Considered a vital component for transporting bulk materials, belt conveyor plays a crucial role in the mining industry. However, its operational state has led to persistent maintenance-related challenges, including significant damage to the equipment, thus requiring a failure analysis of the conveyor components, especially in non-stationary situations due to fluctuations in the material load during the operation of the conveyor [1]. Among those elements, roller has been a central focus of investigations, demanding analyses of operating conditions, reliability evaluations, and data monitoring towards improving accuracy in decision-making, especially regarding such a crucial component [2].

The adoption of predictive maintenance has emerged as a fundamental tool to ensure the integrity of the conveyor system, performed through the monitoring of physical parameters, thus enabling proactive interventions for preventing failures and optimizing

operational reliability [2,3]. Vibration analysis represents one of the main approaches for fault identification; it uses techniques to monitor the signals of rotating machinery and detect any subsequent change in time, frequency, or time-frequency domains. Such a detailed analysis enables diagnosing potential component failures [4].

Due to the non-stationary nature of the signal, conventional techniques (e.g., Fourier transform) show limited efficacy when separately analyzing characteristics in time and frequency domains. Alternative methods such as Wavelet Packet Decomposition (WPD) have emerged as more efficient solutions. WPD addresses vibration signals in the time-frequency domain, effectively analyzing signals with those nonlinear and non-stationary characteristics [5,6]. Due to the complexity of ensuring safe access for inspection and maintenance activities, rollers are often not properly prioritized during periods of scheduled downtime, resulting in their decreased efficiency and possibility of irreversible damage to other systems operations [7].

The advent of Industry 4.0 has promoted more robust communication among remote monitoring systems, hence, a more efficient and proactive management of the integrity of industrial equipment [3,8]. Numerous studies are underway towards diagnosing conveyor belt roller failures through a variety of machine learning algorithms. Machine learning has emerged as an alternative for diagnosing faults through vibration analysis. It enables the creation of models that classify the characteristics of the signal, determining the integrity of the components with high precision [9]. Notable examples of the main machine learning techniques already applied include Support Vector Machine (SVM) [5,10–12].

Towards an effective application of a model, its most significant characteristics (features) must be selected to reduce information redundancy and improve classification performance. Feature selection plays a key role in reducing the dimensionality of the database by preserving a subset of features with greater relevance in the classification process [13].

One of the most used techniques is Principal Component Analysis (PCA), applied for dimensionality reduction in belt fault conveyor classification algorithms [10]. However, Alharbi et al. [9] pointed out that PCA has limitations related to difficulties in multiclass discrimination, particularly in more complex datasets. Kazemi et al. [14] also addressed the limitations of traditional PCA in handling time-varying processes by introducing recursive updates to the correlation matrix. Although the ReliefF algorithm improves upon the original Relief by offering greater robustness to noise and multiclass datasets, it still has limitations. Specifically, it cannot eliminate redundant features and fails to capture conditional dependencies among variables, which may reduce its effectiveness in regression-based or interdependent feature domains, such as fault diagnosis in mechanical systems [15].

Recent advances in artificial intelligence have significantly increased the application of machine learning in fault detection. Vakharia et al. [16] applied Wavelet Packet Decomposition (WPD) combined with Gradient Boosting Decision Trees (GBDT) to diagnose faults in belt conveyor idlers, where the decision tree model inherently performed feature selection by ranking the most informative spectral features. Liu et al. [17] utilized a lightweight attention-based model to detect faults in rotating machinery through acoustic signals, in which the attention mechanism automatically identified the most relevant time–frequency characteristics. In terms of idler fault diagnosis, Muralidharana et al. [18] presented a method based on a decision tree algorithm, which used statistical metrics from idler vibration signals to train the model, achieving good performance in classifying four types of idler faults. Ravikumar et al. [19] also applied decision trees to identify statistical features with a high capacity for diagnosing failures in conveyor belt rollers.

Therefore, alternative feature selection techniques can be explored for fault diagnosis. Two techniques that have been rarely reported in the literature for this purpose are the following: Analysis of Variance (ANOVA), which ranks the main features based on statistical criteria [20], and decision trees, which are commonly used as classification techniques but not often applied as a feature selection method.

This study aims to assess the performance of the decision tree and ANOVA as feature selection methods for diagnosing belt conveyor idlers. Initially, Wavelet Packet Decomposition was applied to vibration signals from conveyor belt rollers for obtaining wavelet energy bands, used as features. The main features (most relevant energy bands) were selected by two different methods (decision tree and ANOVA). Comparison between machine learning models was conducted considering presence or absence of feature selection. SVM, the technique chosen for classification, aimed to identify the model with the best accuracy performance with and without the selection of features.

The paper is structured as follows: Section 2 covers the basic principles of Wavelet Packet Decomposition as a feature extraction method; Section 3 explains fundamentals of Support Vector Machine for classification models; Section 4 describes application of decision tree and ANOVA for feature selection techniques; Section 5 describes the experimental setup and methodology of the study, from data collection to failure diagnosis; Section 6 reports the results and the main findings for multiclass learning, as well as comparative of accuracies achieved by the confusion matrix in different feature selection methods; finally, Section 7 provides the main conclusions and suggestions for future studies.

2. Wavelet Packet Decomposition

The wavelet transform applies to base wavelet, analogously to the sinusoidal functions applied to Fourier transform. The main difference is the shape of sine waves, which are periodic functions with constant amplitude beyond the domain, whereas base wavelets are short-time periodicals with zero linear value outside the domain [21].

The wavelet functions family is determined from the frequency domain, which can contract and expand the wavelet mother by means of the dilation and translation parameter for displaying high- and low-frequency characteristics in any time interval of the signal, respectively. To the end of avoiding data redundancy and calculations on all possible scales, the expansion and translation parameters can be discretized so that the signal analysis remains efficiently accurate. This process is known as Discrete Wavelet Transform (DWT) [22].

According to Rhif et al. [23], DWT is an analysis of non-stationary signals for the detection of structures of spatial and/or temporal domains and extraction of information through frequency variations. During its implementation, Multiresolution Analysis (ARM) was introduced for adapting discrete-time signals of finite length [16]. However, due to restrictions on the functions allowed for MRA, an alternative is the adoption of Daubechies wavelet family of functions [24].

Despite the flexibility of DWT's resolution properties, one of the drawbacks is the poor resolution and discrimination between components of signals at higher frequencies. The creation of Wavelet Packet Decomposition (WPD) from the generalization of wavelet bases has emerged as alternative bases that inherit properties of orthonormality and time-frequency localization to the corresponding wavelet functions [5]. WPD offers a richer analysis based on the decomposition of signals by digital filtering in all frequency bands, creating a set of frequency sub-bands applied for better discrimination of components in the entire frequency domain [25].

The basic principle of WPD is the decomposition of the signal from the transform into low- to high-frequency bands, for which the energy of the spectrum was extracted. Energies

from different frequency bands of a vibration signal can be used as features for enabling fault identification by an intelligent classifier algorithm. Equations (1)–(3) simplify wavelet function (W), wavelet coefficients (w), and band energy (E), respectively. The variable n represents the decomposition level, j is the scaling factor, k is the translational factor, and $f(t)$ is the time-domain sign [5].

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

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

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

3. Support Vector Machine

Machine learning is the science that studies algorithms and statistical models towards computer systems performing certain activities with no programming with explicit commands. Several algorithms can be applied to machine learning, and each of them is more effectively suited to solving a given problem [26]. Regarding classification algorithms, techniques such as Support Vector Machine can be applied.

Considered one of the main tools in science and industry, SVM is one of the pillars of artificial intelligence, due to its effectiveness in classifying a given application in relation to other techniques [27]. It is a machine learning technique that creates a hyperplane with optimal separation of input vectors nonlinearly mapped in a high-dimensional Z feature space. Margins of maximum distance between the nearest vectors are constructed so that the optimal hyperplane can ensure a good generalization of the classes. Therefore, models that classify linearly separable and non-separable data can be generalized. Margins are created from a small piece of data called Support Vectors [28].

A greater variety of decision surfaces, including nonlinear ones, can be constructed for solving the problem of nonlinearity. A decision function can, therefore, be created from the inner product [28]. Regarding nonlinear decision surfaces, the inner product convolution shows variations, whereas for Radial Base Functions (RBF), the decision function $f(x)$ is defined by Equation (4), where α_i is a Lagrangian multiplier, and b is a linear coefficient. $Ky(|x - x_i|)$ is a non-negative function width parameter, expressed by Equation (5).

$$f(x) = sign\left(\sum_{i=1}^{N} \propto_i K_\gamma(|x - x_i|) - b\right), \tag{4}$$

$$K_\gamma(|x - x_i|) = exp\left\{-\gamma |x - x_i|^2\right\}. \tag{5}$$

In general, different types of decision functions can be mapped when the K formula, also known as Kernel, is known. Such functions are very useful for dealing with cases of linearly non-separable data [29].

4. Feature Selection

Since machine learning involves the most diverse techniques, dealing with different approaches requires analyses of the features that will effectively contribute to a good learning model. Therefore, dimensionality reduction has arisen for representing high-dimensional data, without harming their structure, in low-dimensional spaces, with the inclusion of outliers and clusters, enabling graphically visualizing high-dimensional data. The main dimensionality reduction classes are summarized as linear and nonlinear ones, as well as dimensionality reduction algorithms that implement both classes as feature selection [30].

Regarding feature selection, some techniques aim to identify the most significant features, to the detriment of noisier ones that compromise the algorithms' learning process [31]. Among the techniques already applied are the selection of features such as decision tree and ANOVA.

4.1. Decision Tree

Decision tree is one of the most applied tools for machine learning, especially in classification problems, being addressed in areas that require the study of machines, pattern recognition, and statistics [32]. Basically, a partition of the database-driven sample space is created in the form of a tree and non-parametric methodologies are considered, with results interpreted through partitioning called nodes [33].

Initially, each partitioning defines a condition on the distribution of datasets by classes and branches the data samples to other nodes, which will be successively conditioned to other nodes, according to the parameters established in the learning. The last nodes of the tree are called leaves and correspond to predictions [34]. One of the main advantages is the easy interpretation of predictions, based on graphical resources, which can be applied to large scales of data [35]. One of the most well-known decision tree configurations is Classification and Regression Tree (CART) [32].

Once a decision tree algorithm (for feature selection) has been chosen, a set of features is selected for a learning model. Therefore, the most relevant features for a classification model can be visualized in the tree diagram and the least significant ones can be eliminated. Learning techniques in the later stages will be more accurate with a less robust set of information [18,19].

4.2. Analysis of Variance

ANOVA is an excellent technique for statistical tests applied to rejecting null hypothesis H_0, stated as H_0: $\mu_1 = \mu_2 = \cdot = \mu_K$, where μ_i, $i = 1, 2, \ldots, K$ are the means for different classes, identifying the separability between classes. In addition to comparing class averages, ANOVA compares them according to different levels of factors. As an auxiliary tool, *F*-test, represented by Equation (6), is applied [20].

$$F = \frac{\frac{\sum_{i=1}^{K}(Y_i - Y)^2}{K-1}}{\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - Y)^2}{N-K}} \tag{6}$$

where Y_i represents the sample mean in the ith group, n_i is the number of samples in the ith group, Y denotes the overall average of the samples, Y_{ij} it is the jth sample in the ith group, and K represents the number of groups. Therefore, F must be high for rejecting H_0. However, the calculation of F alone does not create a good interpretation for the ranking of features, and another reference value such as p-value is required. p-value is the probability of F being high enough in the observed analysis so that hypothesis H_0 can be rejected. In general, it is rejected if H_0, $p \leq 0.05$ [20].

5. Experimental Setup

The methodology was developed through the following stages (represented in Figure 1): Fault manufacturing, Data collection and processing, Feature extraction and selection, and Classification. Each stage is explained in detail in the following subsections. The equipment used for the application of the vibration analysis and failure classification study was a belt conveyor test rig, operating with an angle of inclination of eight-degree angle and 90 rpm speed, controlled by a frequency inverter.

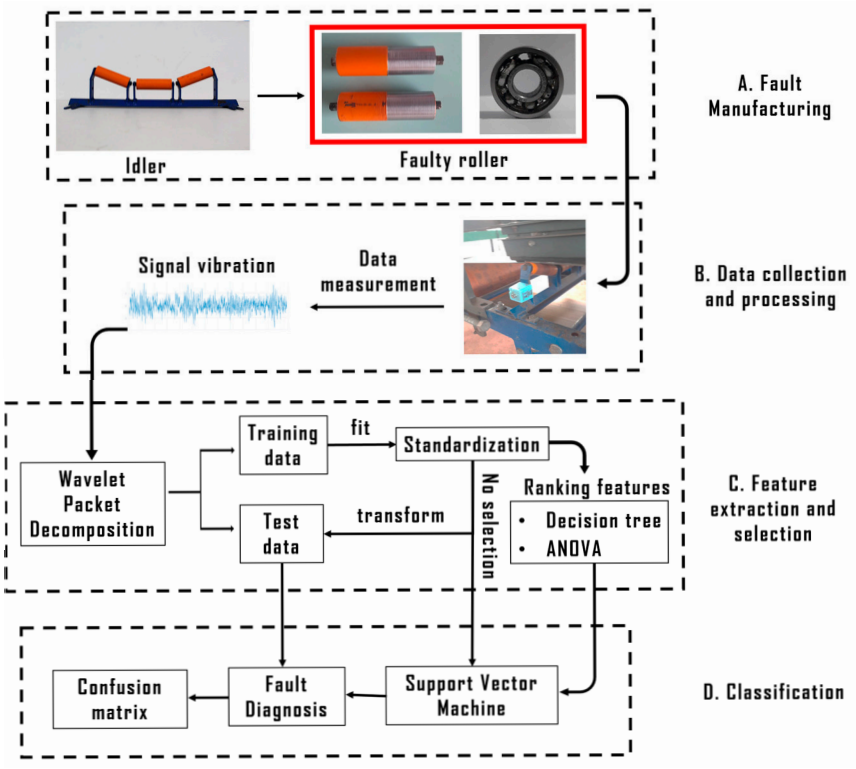


Figure 1. Stages for fault diagnoses.

5.1. Fault Manufacturing

The creation of roller defects was based on simulations in two of the main modes of roller failure, namely, shell surface wear and bearing defects. Initially, artificial defects were implemented on the surface of the rollers with two different severity levels by lathe machining. Enabling a margin of control of defects, specific wear modes in rolls were analyzed from the steps performed in the process of face plating of the roller surface. Two rollers were machined—one with 0.5 mm wear (defined as grade 1) and the other with 1 mm wear (defined as grade 2), as shown in Figure 2.

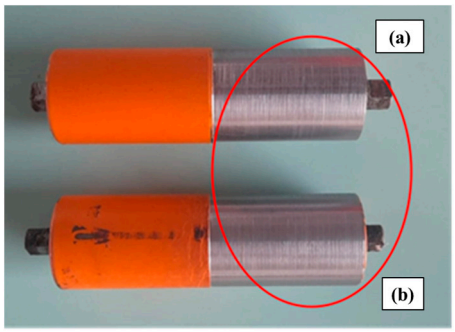


Figure 2. Artificially worn rollers (a) 0.5 mm wear (b) 1 mm wear.

Regarding the manufacture of artificial defects in the roller bearings, the roller was disassembled, and the bearings were removed. Two levels of defects simulated two different severity levels of the defect, where a hole was made for breaking the cage by a hammer drill with a 2.25 mm diameter. In the first severity level, only one of the roller bearings (defined as grade 1) failed, whereas in the second, a hole was drilled in each of the two bearings (defined as grade 2). Figure 3 shows a comparison between a healthy bearing and a bearing with an artificial defect in the cage. The creation of defects in the roller bearing cage is an adaptation of the failure induction from the literature [2].

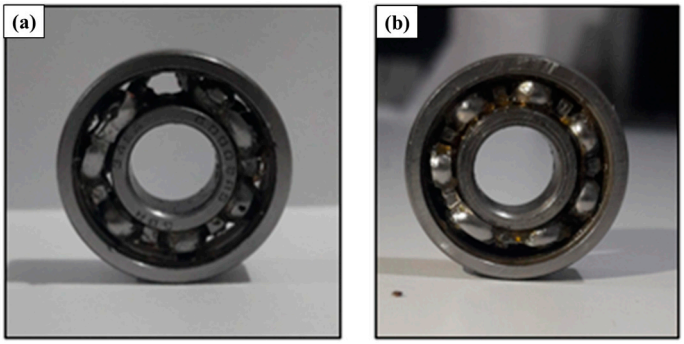


Figure 3. Comparison between (**a**) defective bearings and (**b**) healthy bearings.

5.2. Data Aquisition and Processing

During the experimental procedure, a sensor was installed on the side of the idler frame to measure the vibration of the rollers under each health condition (Figure 4). In addition, both the sensor and the faulty rollers were positioned on different sides of the idler frame to ensure greater signal variability and a broader range of class labels.

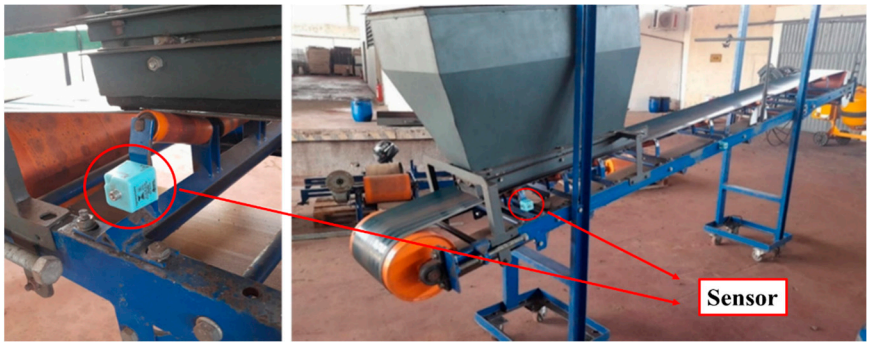


Figure 4. As-built belt conveyor bench.

The sensor selected for the monitoring of the rollers was DynaLogger HF+ model from Dynamox® S.A manufacturer (Florianópolis, Brazil). The basic characteristics of the experimental setup for vibration signals acquisition (Figure 5) are shown in Table 1.

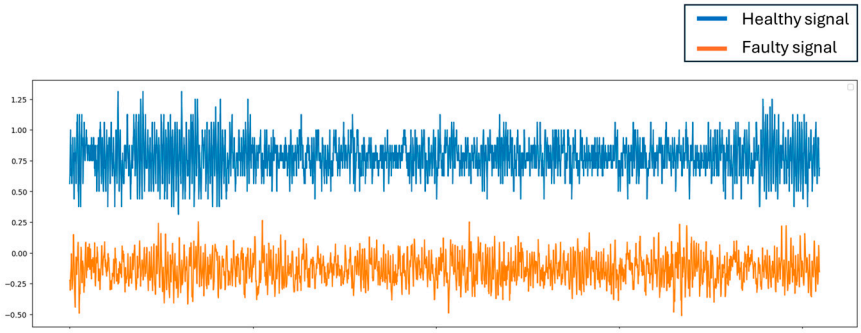


Figure 5. Graphical representation of signals.

Table 1. Experimental Setup Configuration.

Configuration Parameters	
Dynamic range	8 g
Axis	Vertical
Number of samples	4095
Sampling rate	13,151.4 Hz

5.3. Feature Extraction and Selection

From the creation of a database of vibration acceleration in the time domain, the process of decomposition of the signals into energy bands was conducted by WPD. Fifteen levels of decomposition were configured, thus enabling the formation of 16 frequency bands for each sample ($E1, \ldots, E16$). Initially, with the selection of member 'db 8', extracted from Daubechies family, coefficients and wavelet energy for each band were calculated as the selected wavelet function, represented, respectively, by Equations (2) and (3) presented.

The wavelet energy bands were pre-processed so that the training samples could fit into new values within a standardized range. Once the normalized wavelet energy had been calculated, a new database was extracted with each energy band representing a feature of the sample. For each sample, class labels were used to categorize the signal states as either normal or faulty, considering both failure modes and severity levels. Additionally, the algorithm was able to identify faults in rollers positioned on the opposite side of the stand relative to the accelerometer, allowing for the detection of lateral roller failures using only a single accelerometer per idler frame.

Furthermore, an undersampling step was applied, retaining 75% of the data to prevent overfitting during fault classification. Features were handled through feature selection with the decision tree. The decision tree algorithm was then applied from the database with the features obtained. The configuration of the parameters chosen for the decision tree algorithm is shown in Table 2.

Table 2. Parameters for Decision Tree.

Parameters	Status
Criterion (Split Quality Function)	Gini
Maximum depth	4
Minimum number of samples for splitting an internal node	2
Minimum number of samples to be contained in one sheet	1

Figure 6 illustrates the rules established by the model, from features created by the energy bands for classification of the roller conditions. The most significant features chosen in the rules by the model were selected for the application of the techniques that create machine learning models for fault diagnosis.

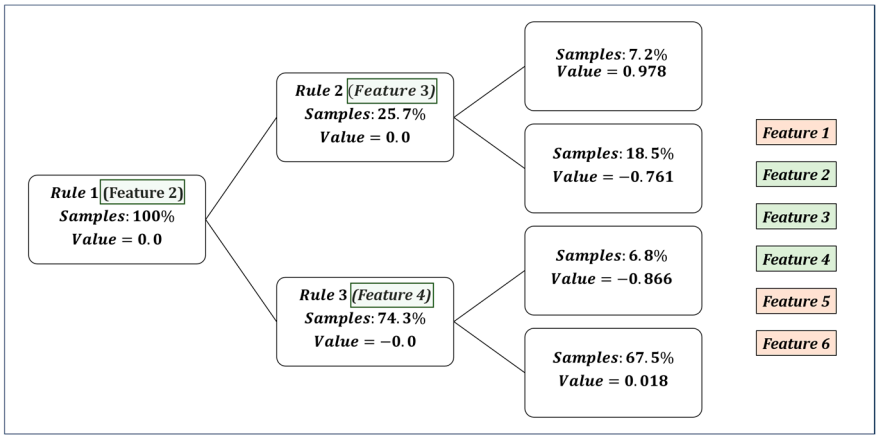


Figure 6. Illustrative feature selection using the Decision Tree model.

After a proper formatting of the new database with the normalized wavelet energy, the features were also selected by ANOVA for comparing averages among the nine different labels of the roller signals. Towards determining statistically significant differences, *F*-test was calculated so that the variability of the data within the class and between classes could

be understood. The *p*-value, which indicates the probability of a class being separable, was also calculated (as represented in Figure 7).

Rank	Features	p-Value
1	Feature 2	0.0012
2	Feature 3	0.0059
3	Feature 4	0.0100
4	Feature 1	0.0199
5	Feature 5	0.0347
6	Feature 6	0.0752



Figure 7. ANOVA for Feature Selection.

5.4. Classification

After the extraction and pre-analysis of features, the six most significant features were selected for comparisons of the diagnostic models with and without them. Data balancing was previously applied for balancing the quantity of data per class for data training towards improving the classification model to be used. The samples were divided into a 75/25% ratio for training and test data, respectively. Otherwise, data normalization was performed after the train-test split step, to avoid data leakage.

SVM was adopted as a learning technique, and algorithms were created for classifying the vibration signals in the vertical direction captured by the accelerometer, with and without the selection of features. A grid search step for hyperparameter optimization was realized based on the variation in main hyperparameters, as shown in Table 3. Learning models performed diagnoses of the rollers, considering detection of failure, failure mode, severity level, and faulty roller position.

Table 3. Configuration of learning algorithm parameters.

Parameters	Status
Learning Technique	SVM
C	1–200
Kernel	Radius Basis Function (RBF)
Gamma (γ)	Scale

6. Results and Discussion

The selection of features for fault diagnosis by decision tree and ANOVA showed some similarities between the techniques. Among the 16 wavelet energy bands, the 7 most significant features for each technique selected as the 1st, 10th, 11th, 12th, 14th, and 16th band energies were best ranked, with only one divergent feature for both feature selection techniques.

Regarding decision tree, in addition to the wavelet energies ranked, the 15th energy was also among the seven most significant features, as illustrated in Figure 8. The root node of the tree displays the sampling distribution rule based on the 1st band energy, which is the feature of highest relevance for classification. The other features adjust the distribution among the nine classes and can assist in the identification of signal characteristics such as type of failure mode, severity level, and roll positioning.

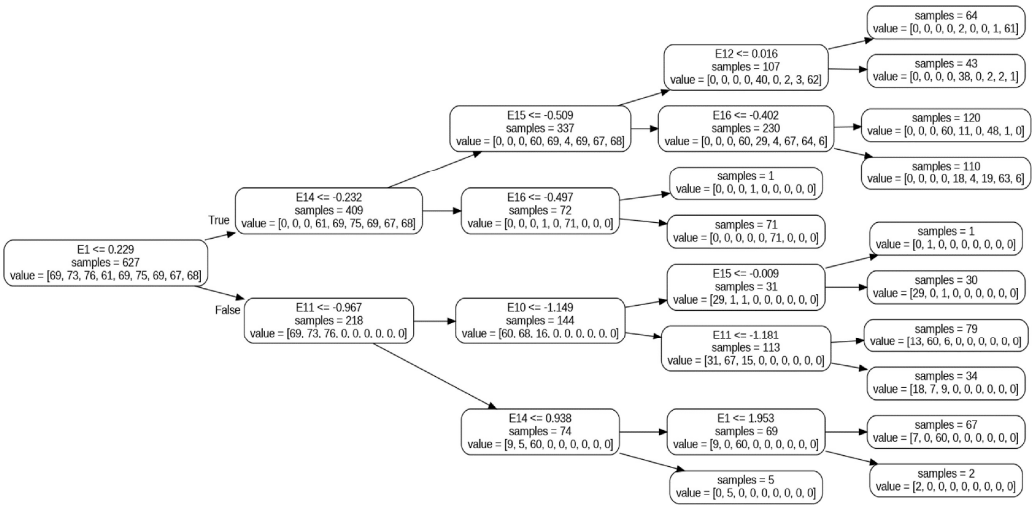


Figure 8. Decision Tree.

According to ANOVA, unlike the decision tree, the 13th energy was ranked among the 7 most significant features. Another difference is the way features are ranked, since the decision tree shows only the most significant ones, eliminating those least relevant, whereas ANOVA ranks all band energies (see Table 4). On the other hand, a similarity between techniques that can be highlighted is the selection of the 1st energy as the most relevant. However, all features had a p-value below 0.05, indicating they are highly differentiable features, and implying that learning can be performed without feature selection, even though it shows a classification accuracy loss.

Table 4. Feature Ranking (ANOVA).

Feature	F-Test	p-Value
E1	1762.193	0
E11	230.5996	6.15×10^{-180}
E13	155.024	2.77×10^{-142}
E14	76.53178	2.52×10^{-87}
E10	59.99508	3.17×10^{-72}
E12	56.13177	1.92×10^{-68}
E16	55.8949	3.31×10^{-68}
E9	46.65241	1.08×10^{-58}
E15	39.38345	1.01×10^{-50}
E4	10.58034	5.60×10^{-14}
E6	10.34173	1.22×10^{-13}
E7	9.24195	4.54×10^{-12}
E8	9.065116	8.13×10^{-12}
E5	8.963274	1.14×10^{-11}
E2	7.667265	8.15×10^{-10}
E3	6.938028	8.99×10^{-9}

After selecting the seven most expressive features for fault diagnosis, SVM learning models were created with a variation in hyperparameter C in three cases, i.e., in a learning algorithm without feature selection, with the application of decision tree, and, finally, with the application of ANOVA. The best models were in the algorithms with the selection of features, implying a decrease in the amount of band energies for the learning of the diagnostic algorithm, reducing the noise that hampers the identification of each class. A confusion matrix was created for the best model in each case towards a better comparison between the predicted class and the actual one.

Figure 9 shows the confusion matrix of the SVM classifier model with no reduction in wavelet energy bands. The main diagonal shows the number of correct answers in each class, and the incorrect diagnoses of the predictor class in relation to the true class are presented. The detection of healthy signals showed no confusion in relation to faulty ones, except rollers with grade 1 surface wear condition, i.e., with wear levels in early stages, which implies a low number of errors due to false negatives.

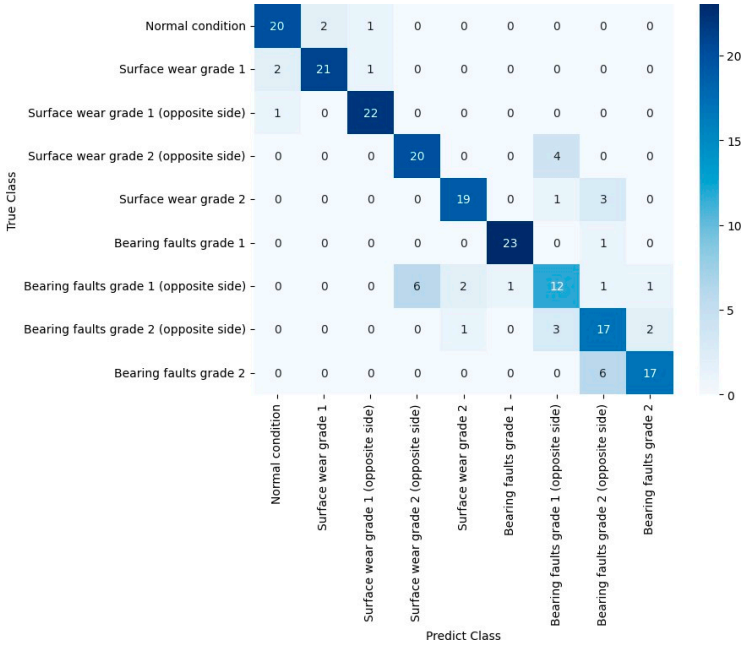


Figure 9. Confusion Matrix (No Feature Selection).

Figure 10 displays the SVM classifier confusion matrix after selecting features by decision tree. Regarding false negatives, a similar behavior was exhibited in relation to the model with no feature reduction. However, the increase in the number of correct answers of the classifier model is remarkable, especially in relation to the classes of signals that showed defects in bearings.

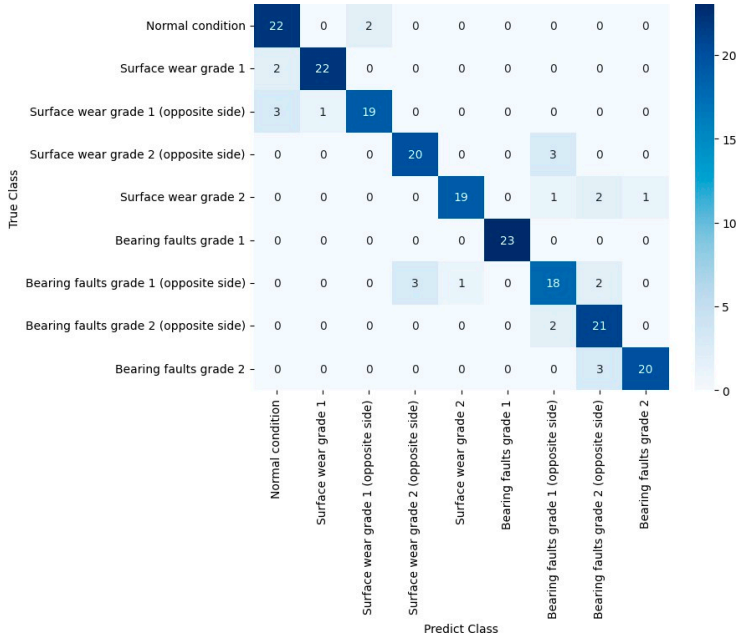


Figure 10. Confusion Matrix (Decision Tree).

Figure 11 shows the confusion matrix of the SVM model with feature selection by ANOVA. In addition to the similarity of the behavior of the model in relation to the other classifier models, the rate of correct answers increased in the diagnosis of the rollers in comparison to the model with no reduction, despite a decrease in accuracy in relation to the model with selection of features by decision tree.

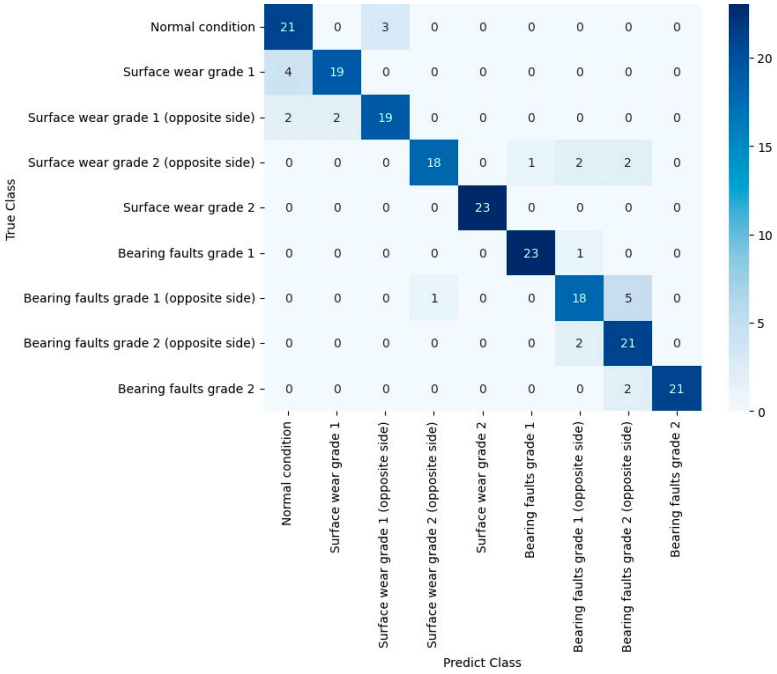


Figure 11. Confusion Matrix (ANOVA).

Some indicators of the diagnosis of roller failures were evaluated from the confusion matrices, as shown in Table 5. Despite the lowest false negative error rate of the model without feature selection, other indicators are superior in models with reduction in features, such as failure mode and severity level. Moreover, the selection of features promoted a more accurate identification of the defective roller on the idler frame.

Table 5. Fault Diagnosis Indicators.

Accuracy	No Feature Selection	Decision Tree	ANOVA
False negative (%)	13.0	18.5	22.2
Roller position (%)	94.0	97.7	97.0
Severity level (%)	96.4	97.8	96.2
Failure mode (%)	90.7	93.9	96.7

7. Conclusions

In general, the machine learning models for diagnoses of failure in belt conveyor rollers have been improved by different feature selection techniques. Wavelet band energies were extracted from vibration signals measured on the idler frame for evaluations of the conditions of the rollers. From the extraction of features, decision tree and ANOVA were applied for their ranking, indicating divergence in only one of the seven most significant features and the main energy bands in the classification of the state of the rolls.

After the selection of the most significant features, different SVM classifier models were compared with and without feature selection techniques. The model applying a decision tree for the ranking of features showed the best indicators such as faulty roller position (97.7%), severity level (97.8%), and failure mode (93.9%). Therefore, applying

feature selection can reduce computational costs without statistically compromising the performance of roller fault classification models. However, in cases involving false negatives, models without feature reduction achieved lower error rates (13%), as shown in Table 5. This suggests that some wavelet band energies discarded during the feature ranking process may carry relevant information for distinguishing between healthy and faulty signals. In all models, errors due to false negatives were detected in the classification between healthy signals and defective signals with grade 1 superficial wear. Therefore, some signals evaluated with levels of wear in early stages, even if in low percentages, may not be detected by diagnostic models.

Towards future improvements, the diagnosis of rollers positioned in the central region of the stand should be included. Moreover, future work should consider applying different loads to the conveyor bench, evaluating the transferability of the trained model to other rigs or operational settings, and comparisons with real field signals should be considered to enhance the practical applicability of classification models and comparisons with traditional feature selection techniques.

Author Contributions: Conceptualization, J.L.L.S., T.B.C., G.S.d.N., J.M.S.d.F. and A.L.A.M. (Alexandre L. A. Mesquita); methodology, J.L.L.S., T.B.C., G.S.d.N., J.M.S.d.F. and A.L.A.M. (Alexandre L. A. Mesquita); formal analysis, J.L.L.S., W.S.S., D.S.B., A.L.A.M. (André L. A. Mesquita) and A.L.A.M. (Alexandre L. A. Mesquita); investigation, J.L.L.S., T.B.C., W.S.S., A.L.A.M. (André L. A. Mesquita) and A.L.A.M. (Alexandre L. A. Mesquita); writing—original draft preparation, J.L.L.S., T.B.C., G.S.d.N., J.M.S.d.F. and A.L.A.M. (Alexandre L. A. Mesquita); writing—review and editing, J.L.L.S. and A.L.A.M. (Alexandre L. A. Mesquita); visualization, J.L.L.S., G.S.d.N., T.B.C. and W.S.S.; supervision, W.S.S., D.S.B., A.L.A.M. (André L. A. Mesquita) and A.L.A.M. (Alexandre L. A. Mesquita). All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded partially by the National Academic Cooperation Program in the Amazon (PROCAD/Amazon—Ref. 88887.473095/2019-00) of the Coordination for the Improvement of Higher Education Personnel of the Brazilian Government (CAPES/Brazil).

Data Availability Statement: The authors confirm that the material supporting the findings of this research is available within the article. The data collected will be made available on request.

Acknowledgments: The authors acknowledge the Dynamox®, the National Academic Cooperation Program in the Amazon (PROCAD/Amazon) of the Coordination for the Improvement of Higher Education Personnel of the Brazilian Government (CAPES/Brazil), and the Federal University of Pará (UFPA), which have significantly contributed to the success of this project.

Conflicts of Interest: Authors Jullyane M. S. de Figueiredo and Danilo S. Braga are employed by the company Dynamox. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of Variance
WPD	Wavelet Packet Decomposition
SVM	Support Vector Machine
DWT	Discrete Wavelet Transform
CART	Classification and Regression Tree

References

1. Lodewijks, G. Two decades dynamics of belt conveyor systems. *Bulk. Solids Handl.* **2002**, *22*, 124–132.
2. Liu, X.; He, D.; Lodewijks, G.; Pang, Y.; Mei, J. Integrated decision making for predictive maintenance of belt conveyor systems. *Reliab. Eng. Syst. Saf.* **2019**, *188*, 347–351. [CrossRef]

3. Susto, G.A.; Schirru, A.; Pampuri, S.; McLoone, S.; Beghi, A. Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Trans. Ind. Inform.* **2014**, *11*, 812–820. [CrossRef]
4. Popescu, T.D.; Aiordachioaie, D.; Culea-Florescu, A. Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: An overview. *Int. J. Adv. Manuf. Technol.* **2022**, *118*, 2883–2899. [CrossRef]
5. Li, W.; Wang, Z.; Zhu, Z.; Zhou, G.; Chen, G. Design of online monitoring and fault diagnosis system for belt conveyors based on wavelet packet decomposition and support vector machine. *Adv. Mech. Eng.* **2013**, *5*, 797183. [CrossRef]
6. Soares, J.L.; Costa, T.B.; Moura, L.S.; Sousa, W.S.; Mesquita, A.L.; Mesquita, A.L.; de Figueiredo, J.M.; Braga, D.S. Fault diagnosis of belt conveyor idlers based on gradient boosting decision tree. *Int. J. Adv. Manuf. Technol.* **2024**, *132*, 3479–3488. [CrossRef]
7. Swinderman, R.T.; Marti, A.D.; Marshall, D. *Foundations for Conveyor Safety: The Global Best Practices Resource for Safer Bulk Material Handling*; Martin Engineering Company: Neponset, IL, USA, 2016. Available online: https://static.martin-eng.com/www.martin-eng.de/download/FoundationsForConveyorSafetyBook.pdf (accessed on 20 April 2025).
8. Muñiz, R.; Nuño, F.; Díaz, J.; González, M.J.; Prieto, M.; Menéndez, Ó. Real-time monitoring solution with vibration analysis for industry 4.0 ventilation systems. *J. Supercomput.* **2023**, *79*, 6203–6227. [CrossRef]
9. Alharbi, F.; Luo, S.; Zhang, H.; Shaukat, K.; Yang, G.; Wheeler, C.A.; Chen, Z. A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models. *Sensors* **2023**, *23*, 1902. [CrossRef] [PubMed]
10. Yang, M.; Zhou, W.; Song, T. Audio-based fault diagnosis for belt conveyor rollers. *Neurocomputing* **2020**, *397*, 447–456. [CrossRef]
11. Roos, W.A.; Heyns, P.S. In-belt vibration monitoring of conveyor belt idler bearings by using wavelet package decomposition and artificial intelligence. *Int. J. Min. Miner. Eng.* **2021**, *12*, 48–66. [CrossRef]
12. Lobato, T.H.; da Silva, R.R.; da Costa, E.S.; Mesquita, A.L. An Integrated Approach to Rotating Machinery Fault Diagnosis Using, EEMD, SVM, and Augmented Data. *J. Vib. Eng. Technol.* **2020**, *8*, 403–408. [CrossRef]
13. Rauber, T.W.; Boldt, F.A.; Varejao, F.M. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2014**, *62*, 637–646. [CrossRef]
14. Kazemi, P.; Armin, M.; Philip, M. Fault detection and isolation for time-varying processes using neural-based principal component analysis. *Processes* **2024**, *12*, 1218. [CrossRef]
15. Subbiah, S.; Jayakumar, C. Opportunities and Challenges of Feature Selection Methods for High Dimensional Data: A Review. *Ing. Syst. Inf.* **2021**, *26*, 67. [CrossRef]
16. Vakharia, V.; Vijay, K.; Pavan, K. A comparison of feature ranking techniques for fault diagnosis of ball bearing. *Soft Comput.* **2016**, *20*, 1601–1619. [CrossRef]
17. Liu, Y.; Miao, C.; Li, C.; Ji, J.; Meng, D.; Wang, Y. A dynamic self-attention-based fault diagnosis method for belt conveyor idlers. *Machines* **2023**, *11*, 216. [CrossRef]
18. Muralidharan, V.; Ravikumar, S.; Kangasabapathy, H. Condition monitoring of self aligning carrying idler (SAI) in belt-conveyor system using statistical features and decision tree algorithm. *Measurement* **2014**, *58*, 274–279. [CrossRef]
19. Ravikumar, S.; Kanagasabapathy, H.; Muralidharan, V. Multicomponent fault diagnosis of self aligning troughing roller (SATR) in belt conveyor system using decision tree: A statistical approach. *FME Trans.* **2020**, *48*, 364–371. [CrossRef]
20. Pena, M.; Cerrada, M.; Alvarez, X.; Jadán, D.; Lucero, P.; Milton, B.; Guamán, R.; Sánchez, R.V. Feature engineering based on ANOVA, cluster validity assessment and KNN for fault diagnosis in bearings. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3451–3462. [CrossRef]
21. Daubechies, I. *Ten Lectures on Wavelets*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992. [CrossRef]
22. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
23. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Appl. Sci.* **2019**, *9*, 1345. [CrossRef]
24. Rowe, A.C.H.; Abbott, P.C. Daubechies wavelets and mathematica. *Comput. Phys.* **1995**, *9*, 635–648. [CrossRef]
25. Wang, X.; Liu, C.; Bi, F.; Bi, X.; Shao, K. Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension. *Mech. Syst. Signal Process.* **2013**, *41*, 581–597. [CrossRef]
26. Mahesh, B. Machine learning algorithms—A review. *Int. J. Sci. Res.* **2020**, *9*, 381–386. [CrossRef]
27. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019. [CrossRef]
28. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2013. [CrossRef]
29. Boswell, D. *Introduction to Support Vector Machines*; Department of Computer Science and Engineering, University of California: San Diego, CA, USA, 2002. Available online: http://pzs.dstu.dp.ua/DataMining/svm/bibl/IntroToSVM.pdf (accessed on 20 April 2025).
30. Wenskovitch, J.; Crandell, I.; Ramakrishnan, N.; House, L.; North, C. Towards a systematic combination of dimension reduction and clustering in visual analytics. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 131–141. [CrossRef]

31. Chen, C.-H. Feature selection based on compactness and separability: Comparison with filter-based methods. *Comput. Intell.* **2014**, *30*, 636–656. [CrossRef]
32. Jijo, B.T.; Abdulazeez, A.M. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
33. Janikow, C.Z. Fuzzy decision trees: Issues and methods. *IEEE Trans. Syst. Man. Cybern. Part B Cybern.* **1998**, *28*, 1–14. [CrossRef]
34. De Ville, B. Decision trees. *Wiley Interdiscip. Rev. Comput. Stat.* **2013**, *5*, 448–455. [CrossRef]
35. Liu, X.; Pei, D.; Lodewijks, G.; Zhao, Z.; Mei, J. Acoustic signal based fault detection on belt conveyor idlers using machine learning. *Adv. Powder Technol.* **2020**, *31*, 2689–2698. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.