

Failure analysis on a water pump based on a low-cost MEMS accelerometer and machine learning classifiers

Análise de falhas em uma bomba d'água baseado em um acelerômetro MEMS de baixo custo e classificadores de aprendizado de máquina

Luciane Agnoletti dos Santos Pedotti¹; Ricardo Mazza Zago²; Jefferson Cutrim Rocha³; José Gilberto Dalfré Filho⁴; Mateus Giesbrecht⁵; Fabiano Fruett⁶

Abstract

This work presents a failure diagnosis tool for a water pump using a low-cost MEMS accelerometer. It was inserted three types of failures: rotor blade (new and damaged), pump soleplate tightness (stiff or loose), and cavitation, in this case on three conditions: none, incipient and severe, totaling twelve fault combinations. These conditions were tested under two different speeds to perform the diagnosis, totaling twenty-four tests. In all cases, the vibration signals from axes X, Y, and Z were acquired. Some features extracted from the vibration spectra from X-axis were used to compose the dataset. These data were analyzed employing logistic regression, a linear support vector machine (SVM), and an artificial neural network multilayer perceptron (ANN-MLP). We compared these three techniques of machine learning and evaluated which one was able to obtain the most accurate result. Using the ANN-MLP, the system was able to detect all three types of failures inserted, with about 100% of accuracy on the rotor blade condition, 92% for anchorage faults, and about 99% accuracy on cavitation state. As a conclusion, it is demonstrated that this classifier algorithm can be used to process the data from the low-cost MEMS accelerometer in predictive maintenance as an accurate tool.

Keywords: MEMS Accelerometer. Diagnosis by vibration. Diagnostic classifiers. Logistic regression. Linear SVM. ANN-MLP.

Resumo

Este trabalho apresenta uma ferramenta de diagnóstico de falhas para uma bomba de água utilizando um acelerômetro MEMS de baixo custo. Foram inseridos três tipos de falhas: lâmina do rotor (nova e danificada), estanqueidade da placa da bomba (rígida ou solta) e cavitação, neste caso em três condições: nenhuma, incipiente e grave, totalizando doze combinações de falhas. Estas condições foram testadas sob duas velocidades diferentes para realizar o diagnóstico, totalizando vinte e quatro testes. Em todos os casos, os sinais de vibração dos eixos X, Y, e Z foram adquiridos. Algumas características extraídas dos espectros de vibração dos eixos X foram utilizadas para compor o conjunto de dados. Estes dados foram analisados empregando regressão logística, uma máquina vetorial de suporte linear (SVM) e perceptron de rede neural artificial multicamadas (ANN-MLP). Comparamos estas três técnicas de aprendizagem da máquina e avaliamos qual delas foi capaz de obter o resultado mais preciso. Usando a ANN-MLP, o sistema foi capaz de detectar todos os três tipos de falhas inseridas, com cerca de 100 % de precisão na condição da pá do rotor, 92 % para falhas de fixação da bomba à base e cerca de 99 % de precisão no estado de cavitação. Como conclusão, é demonstrado que este algoritmo classificador pode ser usado para processar os dados do acelerômetro MEMS de baixo custo em manutenção preditiva como uma ferramenta precisa.

Palavras-chave: Acelerômetro MEMS. Diagnóstico por vibração. Classificadores de diagnóstico. Regressão logística. SVM linear. RNA-MLP.

¹ Profa. Dra., Depto. Acadêmico de Eletrônica, UTFPR, Curitiba, PR, Brasil, E-mail: lucianesantos@utfpr.edu.br

² Me. Engenharia Elétrica, Unicamp, Brasil, E-mail: ricardoz@outlook.com

³ Engenheiro de Aplicação, Brasil, E-mail: jeffersonctr@gmail.com

⁴ Prof. Dr., Faculdade de Eng. Civil, Arquitetura e Urbanismo, UNICAMP, Campinas, Brasil, E-mail: dalfre@unicamp.br

⁵ Prof. Dr., Faculdade de Eng. Elétrica e Computação, UNICAMP, Campinas, Brasil, E-mail: mateus@fee.unicamp.br

⁶ Prof. Dr., Faculdade de Eng. Elétrica e Computação, UNICAMP, Campinas, Brasil, E-mail: fabiano@dsif.fee.unicamp.br

Introduction

Fault diagnosis has been widely used in the industry for several years. The investment in predictive maintenance may represent a decrease in operational costs because it implies a significant drop in the unscheduled machine stops (FLORINO; MELLO; CARAZZAI, 2014). Moreover, modern diagnosis systems reduce the maintenance time, optimizing its planning because those tools can indicate which is the type of incipient failure (JARDINE; LIN; BANJEVIC, 2006; NAKAJIMA, 1988; STEEGE, 1996).

Rotating machines are subject to fatigue, wear, deformations, anchorage base accommodation, among other problems. These phenomena cause shaft misalignment, slack in straps and gears, cracks, unbalance, and many other defects. Those failures increase the vibration, which can be measured and analyzed, and, from that signal, it is possible to identify and in some cases go further: diagnosing the mechanical malfunction (RAO, 2010).

Generally, the vibrations tend to increase along the time, causing vibration on several frequencies with significant amplitudes which may result in machine malfunctions, implying maintenance costs and machine failures. Many standards were created to indicate acceptable vibration thresholds for rotating machines (IEEE..., 2007). These thresholds help evaluate vibration levels, allowing the operators to justify the maintenance decisions (NORTON; KARCZUB, 1994).

The forces exerted on different machine parts introduce specific excitations, which may be used to identify possible failures. Thus, a measurement of the vibration behavior is suitable for determining the core of the problem, allowing a more accurate diagnosis, indicating the potential fault, like a bearing, bushing, gear, etc. The machine vibrations occur in a wide range of the spectrum; thus, it is possible to identify some malfunctions using the corresponding frequencies (BRAUM, 1986).

According to the Hydraulics Institute, 20% of total energy consumption in the industry is provided by fluid moving systems (pumps) (AUGUSTYN, 2012; BAZANINI, 2018; MATTOS; FALCO, 1998). Thus, this type of equipment is essential for the perfect operation of the productive process. A common malfunction on water pumps is the cavitation. This phenomenon is characterized by the small steam bubbles shaping in the middle of the fluid when it is subject to a specific condition of pressure and temperature. When one of these bubbles, also known as cavities, implode, an enormous amount

of energy is released, causing wear on the metal surface blade, including debris detachment, which can propagate to the pipe and compromise other equipment present in the system (WHITE, 2016).

This kind of phenomenon significantly reduces the lifetime of a pump, due to the wear which it causes in its rotor. The cavitation is related to the pipe pressure, among other variables; thus, the pressure measurement is usually employed to analyze and control this phenomenon. The installation of equipment to measure this variable is invasive into the process, requiring pipe cuts, and flanges assembly what leads to an additional cost to monitor the cavitation (KAYA *et al.*, 2008; WHITE, 2016).

However, the cavities rupturing generates high amplitude vibrations, increasing the ambient noise, and propagating for the whole system. The relationship between the cavitation condition and the vibration is governed by a phenomenon that depends on the forces exerted by the rupture of the steam bubbles and on the response of the mechanical structure to these forces. These different cavitation levels result in other vibrations excitations with different frequencies and amplitudes which can be measured and analyzed (WOWK, 1991).

Thus, this vibration signal can be used to classify the cavitation level. Computational tools have been used in this classification task. A typical tool that is mostly used in classification problems is Logistic Regression. In some cases, a non-linear classification tool is used to aid in this diagnosis too, like artificial neural networks (ANN), which are tools capable to approximate any differentiable function (CYBENKO, 1989). In this paper, the classification was made using these three technics: logistic regression, linear support vector machine (SVM), and multi-layer perceptron (MLP).

Any malfunction in a rotating machine may cause changes in its vibration as unbalance, misalignment, cracks, among others. Each machine element induces its excitement, which identifies a specific disorder (RAO, 2010; WOWK, 1991). Considering it, other disturbances besides cavitation were applied to simulate other common malfunctions. These disturbances are related to the mechanical equipment situation. During the tests, two rotors were used, one of them with a set of new blades and the other one with damaged blades, that have already suffered cavitation effects. Another disturbance applied is related to the pump anchorage, which was fixed stiffly and loosely. In all cases, the system was operated at two different speeds: 900 RPM and 1368 RPM.

The relationship between cavitation and vibration in centrifugal pumps has been investigated in some works. In Nasiri, Mahjoob and Vahid-Alizadeh (2011) a cavitation diagnosis based on vibration is presented. The vibration was acquired using three piezoelectric accelerometers, which costs about US\$100.00 per unit, and the data were analyzed with an ANN. No other failures besides cavitation were investigated, and the reference discusses the influence of the position of the sensors in the diagnosis; three-speed are tested, and the first two harmonic frequencies are used as an input feature, and all the data were acquired using a commercial analyzer B&K 3560C.

In Stopa, Filho and Martinez (2014), a method to detect incipient cavitation using Load Torque Signature Analysis (LTSA) correlating with the vibration FFT analysis under different pump speeds is presented. In that work the load torque was estimated measuring the pump current and monitoring the vibration through a not specified sensor; however, an index of precision in the diagnosis was not presented. As in Nasiri; Mahjoob and Vahid-Alizadeh (2011), no other failure besides incipient cavitation was identified.

In Hamond *et al.* (2017) it is presented that it is possible to diagnose faults on centrifugal pumps using low frequencies analysis; The faults investigated were the inlet vane damages and bearing outer race faults. The sensor used to acquire the vibration signal was non-specified. In that work, the modulation signal bispectrum method was used to characterize the vibration signals. This method is employed to extract deterministic characteristics of modulating components in a low-frequency band for diagnosing both the bearing defects and impeller blockages. No investigation about cavitation was presented.

In Dutta *et al.* (2018) a machine learning algorithm is presented to diagnose the cavitation status on centrifugal pumps, being able to detect if there is cavitation or not; incipient cavitation was not investigated. The support vector machine classifier was used in a machine learning algorithm. However, the signals used to characterize the failure were used as the system pressures measurements such as the static head, vapor pressure, gauge pressure, and friction head. Consequently, an invasive method was used to install the sensors.

In Sánchez *et al.* (2018) the vibration signals acquired by an accelerometer were compared with the ISO 10816 standard of vibration severity. The signals were measured under normal operation, cavitation operation (similarly as done in Dutta *et al.* (2018) and under dif-

ferent speeds too. The vibration severity RMS was determined to compare according to ranges to ISO 10816 standard.

In this work, the results for fault diagnosis are not only cavitation but also for incipient cavitation, and two other failures are presented. The introduction of the other failures was done to check the capacity of the classifier algorithm to differentiate the phenomena that cause vibrations and still perform an accurate diagnosis. The data was also collected at different speeds to validate the diagnosis for different operational conditions.

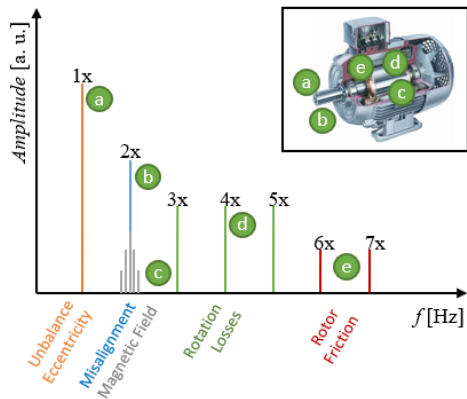
Another contribution of this article is the use of low-cost equipment. A MEMS accelerometer (microelectromechanical systems) NXP MMA8451Q model embedded on the development board FRDM-KL25Z was chosen. This equipment costs less than US\$1.20 per unit (NXP, 2017) and presents low energy consumption. A Matlab interface was developed to acquire the vibration signals (PEDOTTI; ZAGO; FRUETT, 2017). This kind of equipment is easy to install because the process is non-invasive and does not require a system shutdown.

This work is structured as follows: theoretical basis, where the relationship between faults and vibration is presented and is explained how cavitation arises and what are its implications on the equipment; experimental apparatus, where the test plant is presented; data acquisition, where the data collection device is presented; analysis techniques, where the three methods used are explained; operation conditions, explaining which was the measurement conditions on each test; data analysis, presenting how the data was prepared and divided and finally the results.

Theoretical Basis

The rotating machines generally present periodic vibrations (PEDOTTI; ZAGO; FRUETT, 2017; RAO, 2010). A Fourier Series can represent any periodic waveform. The knowledge of the vibration spectrum is essential to diagnose several anomalies because each kind of malfunction presents a specific characteristic signature. For example, the unbalance, affects directly the FFT amplitude corresponding to the machine fundamental frequency (RAO, 2010; WOWK, 1991). Figure 1 shows some cases of malfunctions on rotating machines concerning affected frequency.

Figure 1 – Example of machine malfunctions and their corresponding frequencies.



Source: The authors.

Cavitation on water pumps

In a liquid flow, keeping the temperature constant, bubbles or steam cavities are generated whether the pressure is reduced to the steam pressure of that liquid. Hence the term cavitation. In pumping systems, the liquid transported can vaporize in the pump suction or its rotors, limiting the suction capability and causing other undesirable effects as pressure fluctuation, vibration, additional noise, and erosion.

In industries or utilities, the conservation and maintenance of the centrifugal pumps represent a significant resource of labor and materials (SCHRODER; LUCCA; DALFRÉ FILHO, 2015). Nonconformities, as pumps operating with cavitation, cause essential losses to the productive system. Engineers, technicians, and other professionals involved in the installation and maintenance use the theoretical foundations and experiments performed to comprehend the causes, measure the effects, and to prevent damage caused by cavitation phenomena.

Commonly used quantities to characterize the cavitation are the Net Positive Suction Head (NPSH) or the cavitation index. The NPSH calculation is the most used for the technician involved in pumping system designs (SCHRODER; LUCCA; DALFRÉ FILHO, 2015). In this case, two indexes are calculated, the NPSH available (NPSHa) and NPSH required (NPSHr). The NPSHa refers to the total system pressure measured in the suction pump blind flange in a specific flow and water temperature. The NPSHa is a characteristic of the pumping system and depends on the altitude of the installation, temperature, and the liquid characteristics and also on the hydraulic configuration of the system.

Table 1 – Cavitation Condition × Measured Pressure.

Cavitation Condition	Measured Pressure
None	0.5 psi
Incipient	0.3 psi
Severe	0.1 psi

Source: The authors.

On the other hand, the NPSHr represents the energy of the liquid in the suction flange of the pump, required to reach the rotor, without vaporization. The NPSHr depends on elements of pump design as rotor diameter, engine speed, and specific rotation. Finally, for a given flow, if the NPSHa calculated is greater than NPSHr, provided by the pump manufacturer, the system will not cavitate.

The cavitation index (σ) is a dimensionless variable that relates the energy of pressure available for given kinetic energy which favors or, even, indicates the susceptibility of occurrence of the phenomenon. In Knapp, Daily and Hammit (1970) the cavitation index σ , is defined by equation (1)

$$\sigma = \frac{p_0 - p_v}{\frac{\rho v^2}{2}}, \tag{1}$$

where p_0 is the reference pressure in the flow, p_v the vapor pressure, ρ the specific mass, and v the flow velocity.

High values of σ reflect, in general, the absence of cavitation, while reduced values indicate the presence of cavitation. Despite the ease of calculating the index, its experimental determination is not too easy (SCHRODER; LUCCA; DALFRÉ FILHO, 2015), mainly for incipient cavitation. An experiment, from a flow without cavitation, requires the imposition of pressure reduction by some device and detection of parameters to characterize the presence of the phenomenon. For example, many experiments adopt visual observation to detect the presence of steam bubbles, which when detected, indicate the presence of cavitation and intensity. This requires a transparent device, in this case, in the suction pump, which is usually difficult to install (SCHRODER; LUCCA; DALFRÉ FILHO, 2015; KNAPP; DAILY; HAMMIT, 1970).

In this experiment, some classifiers algorithms are trained to detect the cavitation from vibration signals. A manometer was installed in the pump suction point in order to provide data for the supervised training of those algorithms. The pressures and cavitation conditions are presented in Table 1. This procedure needs to be done only once for each pump type so that it is possible to determine its profile.

Table 2 – Costs of main commercial devices.

Commercial devices	Cost Per Unit
VM220 Portable	US\$ 785
VM-6360 Hand Held	US\$ 380
HS-620 & HS-630	US\$ 1.015
VSA101 & VSE002	US\$ 1.460
FLUKE 805	US\$ 2.200
Tecnoferramentas VM-6380	US\$ 1.090
Omega HHVB82 U	US\$ 1.435
OneProd Falcon SMART	more than US\$ 17.000

Source: Pedotti, (2019).

Through installing access plugs on the piping, manometers can be installed to measure the pressure. The plugs can be sealed later and the manometers can be used at other points in the process. Other techniques for cavitation detection using pressure transmitters can be done. However, pressure transmitters are significantly more expensive when compared to manometers.

Less expensive procedures as the use of vibration sensors mean a significant reduction in installation costs. However, commercial piezoelectric accelerometers are also expensive, making its installation practically unfeasible on some machines.

The accelerometer used is not yet prepared to face the bad weather of industrial environments, such as dust, shock, among others. However, a device can be developed for this purpose using similar technologies that may offer results analog to those obtained in this article. Table 2 gives a summary of the main vibration measuring devices for the industrial environment used in the market and their respective costs. Some of the devices listed in the table have superior technologies. However, there is a market potential to be exploited for more cost-effective machine health monitoring devices.

The use of non-invasive equipment with low-cost and easy installation makes this monitoring feasible, and this can result in a significant reduction in maintenance costs. This is the purpose of the MEMS device shown in section *Data acquisition*.

Experimental apparatus

In this section, the test bench and the data acquisition device are presented.

Test bench

Hydraulic, electrical, and mechanical components that constitute the test bench for the tests with the centrifugal pump are shown in Figure 2.

The water used in the experiment is recirculated from a bottom tank and boosted to the upper reservoir. The upper tank has 8.4 m³. In the center of the tank has a flow-calming system, in the shape of a circular crown, with 2.6 m of external diameter and 2.3 m of internal diameter and 0.85 m high, built-in perforated metal grid, and its interior filled with about hundred thousand 0.02 m diameter glass beads. Thus, the flow which enters the suction pipe of the pump is stabilized. On the center of the circular crown, a suction water intake pump with 0.15 m diameter, and a filter was maintained with a 0.6 m minimal submergence level during the experiment. Hanging, in the center of the tank too, a vortex breaker built with a wood shape prevents the occurrence of swirls over the water intake.

From the water intake of the upper tank, the water is suctioned by a downward pipe with 1.95 m of length and 0.15 m diameter. Upstream the tested pump and the suction line, there are a valve and a pressure transducer, separated by a distance of 1.2 m. A manovacuumeter was installed too to allow a fast visual verification of changes and the operation point tuning. Upstream of the pump flange, a rectilinear section pipe with 1.2 m length from the pressure transducer finishes the suction line.

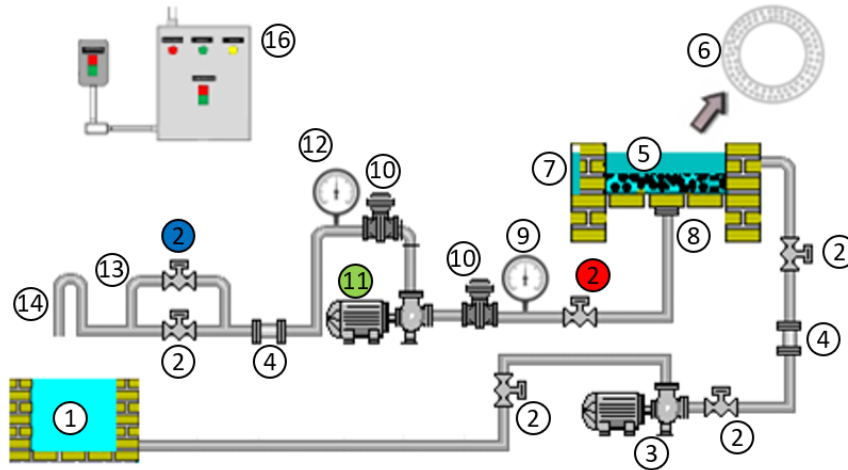
The equipment used during the experiments is a horizontal centrifugal pump Ns 147, mono stage, radially split, fixed to concrete by anchors. The pump and motor assembly are aligned by laser. The frequency inverter is responsible for supplying power to the three-phase electric motor (WEG IP55, 22 kW, 4 poles, 1800 RPM, $\eta = 89\%$) with a manually controlled acceleration ramp.

Data acquisition

The device chosen as a vibration data acquisition tool was the MEMS MMA8551Q accelerometer, embedded on the development board FRDM-KL25Z. This capacitive accelerometer is a triaxial sensor, with 14 bits resolution (NXP, 2018). The three-axis acquisition device was assembled using the configuration shown in Figure. 3.

The acquisition provided by the accelerometer was communicated using the USB (Universal Serial Bus) communication between the acquisition board and a computer with Matlab. Figure 4 shows the block diagram of the device installed in the pump.

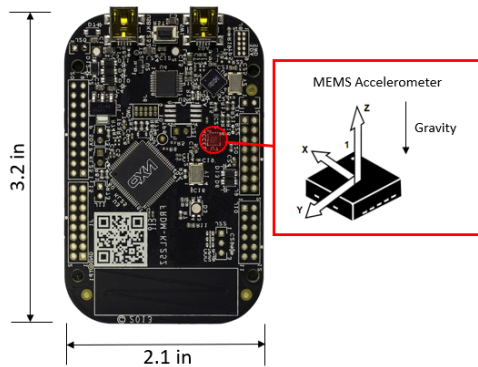
Figure 2 – Test bench diagram. 1. Bottom tank; 2. Valve; 3. Supply pump; 4. Flowmeter; 5. Upper tank; 6. Flow-calming; 7. Level ruler; 8. Water intake; 9. Manovacuumeter; 10. Pressure Transmitter; 11. Centrifugal pump (which is the one under study); 12. Manometer; 13. Bypass; 14. Siphon; 15. Inverter command panel.



Source: The authors.

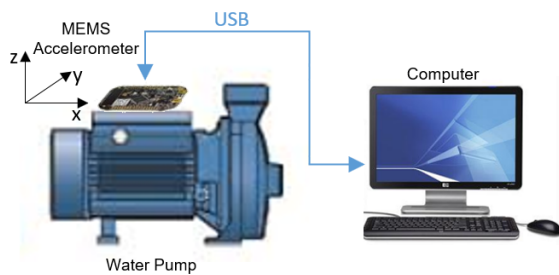
This system was developed by the authors of this paper especially to measure vibrations in rotating machines (PEDOTTI; ZAGO; FRUETT, 2017).

Figure 3 – Acquisition board model FRDM-KL25Z with the MEMS accelerometer MMA8451Q on board and its coordinates.



Source: The authors.

Figure 4 – Block diagram of the experimental apparatus.



Source: The authors.

Analysis techniques

Three known classifiers were implemented to evaluate which analysis technique is suitable to obtain the best classification accuracy. These classifiers are presented in this section.

Logistic regression

Logistic regression can be used for various classification problems. It is one of the most simple and commonly used Machine Learning algorithms for two-class classification. Logistic regression describes and estimates the relationship between a dependent binary variable and independent variables, and predicts the probability of occurrence of a binary event using a logit function (LAROSE, 2006). The model to calculate the probability can be obtained by equation (2)

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}, \quad (2)$$

where p is the probability, β_i are the parameters of the model and x_i , $i = 1 \dots n$ are the values in the observed data.

In logistic regression: the dependent variable follows a Bernoulli Distribution, and the parameter estimation is done through maximum likelihood (HARRELL JUNIOR, 2001).

Linear support vector machine (SVM)

A support vector machine is a discriminative classifier formally defined by a separating hyperplane. The derivation of the hyperplane in linear SVM is done by transforming the problem using linear algebra (ROJO-ÁLVAREZ *et al.*, 2018). This classifier is capable to perform multi-class classification. The LinearSVC is an implementation of support vector classification (SVC) for the case of a linear kernel. For a training dataset with n points of the form $(\mathbf{x}, y_1), \dots, (\mathbf{x}_n, y_n)$ where the y_i are either 1 or -1 , each indicating the class to which the point \mathbf{x}_1 belongs and \mathbf{x}_1 is a p -dimensional real vector. Any hyperplane function is given by equation (3):

$$\mathbf{w} \cdot \mathbf{x} - b = 0, \tag{3}$$

where \mathbf{w} is the normal vector to the hyperplane. The parameter $\frac{b}{\|\mathbf{w}\|}$ determines the offset of the hyperplane from the origin.

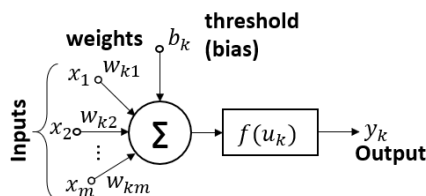
Artificial neural network

Biological neural networks inspire artificial neural networks (ANN). They are general functions approximators and can be applied to almost any machine learning problem. It uses a nonlinear mapping from the input to the output space. Furthermore, it is possible to create conditions that reproduce cognitive ability and processing skills which are much desirable in several applications (HAYKIN, 2008; GOLDBERG; HIRST, 2017).

When a system presents multidimensionality and its variables are subject to nonlinear interactions and when is possible to collect data and learn with that, the neural network becomes one of the possible ways to model that system.

The mathematical modeling of an artificial neuron presents simplifications with the purpose to represent the most significant system aspects. Therefore, less considerable details are discarded, enabling its modeling. Figure 5 shows the mathematical model of an artificial neuron.

Figure 5 – Mathematical model of an artificial neuron.



Source: The authors.

The output y_k of the neuron k can be written by equation (4)

$$y_k = f(u_k) = f\left(\sum_{j=1}^m w_{kj}x_j + b_k\right), \tag{4}$$

where u_k is the input signal of the activation function $f(u_k)$, w_{kj} is the weight associated with the input j , x_j it the j -th input, b_k is a bias and m is the number of inputs.

A simplification can be made in equation (4) including the *bias* as an input signal with value $x_0 = 1$ with associated weight $w_{k0} = b_k$, as shown in equation (5)

$$y_k = f(u_k) = f\left(\sum_{j=0}^m w_{kj}x_j\right) = f(\mathbf{w}^T \mathbf{x}), \tag{5}$$

where $\mathbf{w}^T = [w_{k0} \ w_{k1} \ \dots \ w_{km}]$ and $\mathbf{x} = [x_0 \ x_1 \ \dots \ x_m]^T$. The connection process between the artificial neurons leads the generation of synapses and construction of an ANN (Artificial Neural Network).

Multi-layer perceptron

The Multi-Layer Perceptron (MLP) is an ANN feed-forward class. An MLP consists of, at least, three nodes: an input layer, a hidden layer, and an output layer. Each node is a neuron that uses nonlinear activation functions. In the Multi-Layer Perceptron (MLP), the output from each neuron of the previous layer is the input for every neuron in the next layer. This structure is known as perceptrons neural network with an intermediate layer. The MLP uses a learning technique called backpropagation for training. It can classify data that is nonlinearly separable (HAYKIN, 1998; ROJAS, 1996).

The activation functions used in this work were: Rectified Linear Units (ReLU), equation (6) and Sigmoid, equation (7)

$$f(x) = \max(0, x), \tag{6}$$

$$f(x) = \frac{1}{(1 + e^{-x})}. \tag{7}$$

The ReLU (NAIR; HINTON, 2010) is a general activation function used in most cases currently, and the sigmoid generally works better for classifiers. These activation functions have been chosen due to this characteristics.

Operational conditions

The experiments were performed under twelve operational conditions of the pump, according to the rows of the Table 3. In this table, the column *Target ANN* indicates which value is considered for each condition to train the classifiers. In the array, the first position is used to indicate the rotor blade condition, 0 in case of new and 1 in case of damaged. The second it is used to indicate the base fixation condition, 0 in case of stiff or 1 in case of loose. The third, fourth and fifth positions, indicates the cavitation state being, 1 in case of no cavitation in the third position with 0 in the following two positions, 1 in case of incipient cavitation in the fourth position with 0 in the previous and next positions and 1 in case of severe cavitation in the fifth position with 0 in the two previous positions.

The pump speed was adjusted using a frequency inverter, in operation mode V/f, to 76% (1368 RPM) and 50% (900 RPM) of nominal value, totalizing 24 tests. The accelerometer was fixed in the pump mechanical seal in vertical the position, as shown in Figure 4.

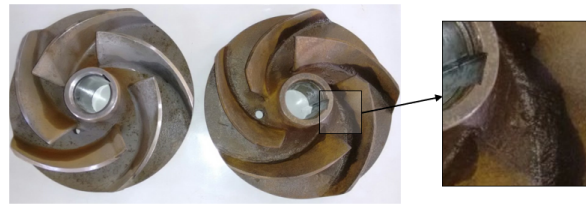
In the operation mode considering the input pressure normal, the pump operates in a steady state without cavitation, in other words, $NPSH_a > NPSH_r$. In this case, the control valve (Figure 2: 2 red) (suction) is completely opened, and the containment valve is closed entirely. Only after starting the motor and adjusting the speed pump (Figure 2: 11 green), the containment valve (Figure 2: 2 blue) is opened. The vibration data was collected after the system is in steady state operation.

In the operation where the input pressure presents incipient cavitation characteristics, the system operates in a mode similar to the normal, except for the reduction in the suction pressure through the closure of the suction control valve, which causes a decrease on the total head gauge and the presence of noise in the rotor, which are markers that classify cavitation.

Each type of cavitation was evaluated using a pressure measuring with a manometer installed in the suction point, with was determinate in 0.5 psi for normal conditions, 0.3 psi for incipient cavitation and less than 0.1 psi for severe cavitation. Thus, from the indication of these markers, the vibration data were acquired.

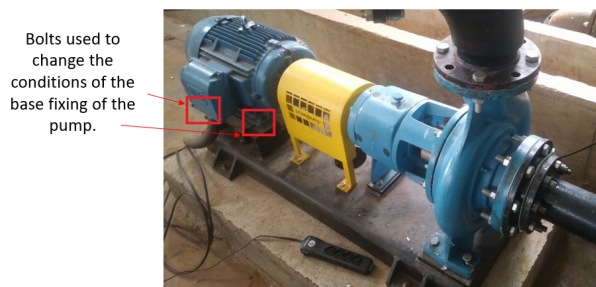
For the tests with the rotor blade, two blades were used, one new and other with severe damage caused by cavitation. Figure 6 shows the rotor blades that were used. For the tests with different base fixation conditions, two fixation bolts were loosened as shown in Figure 7.

Figure 6 – Rotor blades used in the tests: (left) new (right) damage (with erosion highlighted).



Source: The authors.

Figure 7 – Bolts used to change the base fixing conditions.

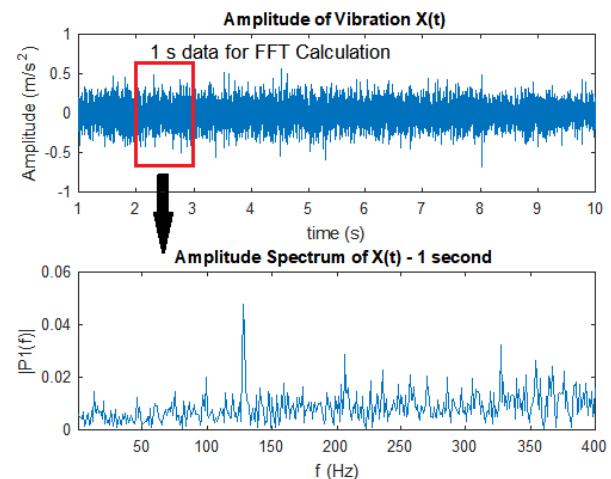


Source: The authors.

Data analysis

The accelerometer can measure the vibration of the three axes over time. Thus, the choice of relevant data to analyze needs to be done. A feature widely used in vibration analysis is the amplitude in the frequency domain. So, the FFT of vibration was computed for data in 1 s intervals, along with the RMS value of X, Y, and Z. The results for one of the intervals for the axis X are shown in Figure 8.

Figure 8 – FFT calculated for computer analysis using *Test 1* data.



Source: The authors.

Table 3 – Combinations of tests performed.

Test	Rotor Blade	Base Fixation	Cavitation	RPM (%)	Binary Target
1	New	Stiff	None	76	[0,0,1,0,0]
2	New	Stiff	Incipient	76	[0,0,0,1,0]
3	New	Stiff	Severe	76	[0,0,0,0,1]
4	New	Loose	None	76	[0,1,1,0,0]
5	New	Loose	Incipient	76	[0,1,0,1,0]
6	New	Loose	Severe	76	[0,1,0,0,1]
7	Damaged	Stiff	None	76	[1,0,1,0,0]
8	Damaged	Stiff	Incipient	76	[1,0,0,1,0]
9	Damaged	Stiff	Severe	76	[1,0,0,0,1]
10	Damaged	Loose	None	76	[1,1,1,0,0]
11	Damaged	Loose	Incipient	76	[1,1,0,1,0]
12	Damaged	Loose	Severe	76	[1,1,0,0,1]
13	New	Stiff	None	50	[0,0,1,0,0]
14	New	Stiff	Incipient	50	[0,0,0,1,0]
15	New	Stiff	Severe	50	[0,0,0,0,1]
16	New	Loose	None	50	[0,1,1,0,0]
17	New	Loose	Incipient	50	[0,1,0,1,0]
18	New	Loose	Severe	50	[0,1,0,0,1]
19	Damaged	Stiff	None	50	[1,0,1,0,0]
20	Damaged	Stiff	Incipient	50	[1,0,0,1,0]
21	Damaged	Stiff	Severe	50	[1,0,0,0,1]
22	Damaged	Loose	None	50	[1,1,1,0,0]
23	Damaged	Loose	Incipient	50	[1,1,0,1,0]
24	Damaged	Loose	Severe	50	[1,1,0,0,1]

Source: The authors.

The sample rate of the accelerometer is 817Hz. Thus the FFT has 408 samples, with 1Hz of resolution. During the preliminary data analysis, it was observed that only the spectral amplitudes for vibrations in X-axis, the RMS values for vibrations in three axes, and the speed provided enough data to classify the operational conditions studied in this work. The total time measured for all tests was 2278 seconds (about 1,58 s/test). So, the data array has 2278 lines.

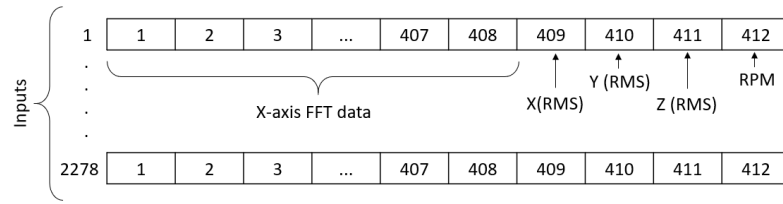
Two different sets of features, each one of them including the X-axis spectrum, RMS values for the three axes, and speed, were considered. In the first one, all the 408 components calculated for integer frequencies of the X-axis spectrum were considered, totalizing 412 features. This dataset was called *dataset A*.

However, it is well known to be possible to detect the majority of low-frequencies faults observing only the main harmonics amplitude (PEDOTTI; ZAGO; FRUETT, 2017). Thus, an analysis using these features was made to verify if these were enough to diagnose the same faults.

So, using the same quantity of data but with fewer features, the same methods were tested. Therefore, in this dataset array, there are 2278 lines, with 14 features each, as presented in Figure 10. This dataset was called *dataset B*.

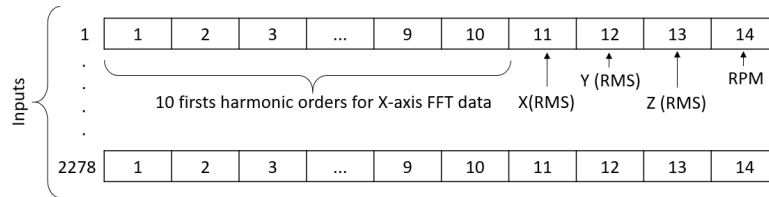
For each result, a ROC (Receiver Operating Characteristic) curve was plotted as a metric to evaluate the classifier output quality. The ROC curves feature a true positive rate on the Y-axis and false positive rate on the X-axis. Thus, the top left corner of the plot is the *ideal* point; a false positive rate of zero, and a true positive rate of one. Then a curve is defined from the point (0,0) to the point that characterizes the true positive rate and the false-positive rate for a given classifier and from that point to the point (1, true positive rate). A larger area under that curve is usually better (POWERS, 2011). To extend the ROC curve and area to multi-class classification, binary targets were chosen. The AUC (area under the ROC curve) measures the entire two-dimensional area underneath the entire ROC curve and provides an aggregate measure of performance across all possible classification thresholds.

Figure 9 – Features extraction from vibration data to use in the computer analysis (whole spectrum): *dataset A*.



Source: The authors.

Figure 10 – Features extraction from vibration data to use in the computer analysis (10th firsts harmonics orders): *dataset B*.



Source: The authors.

The AUC can be interpreted as the probability that the model ranks a random positive sample more likely than a random negative sample.

The programming language Python was chosen to train the classifiers (SCIKIT-LEARN, 2018). The Keras library was used (KERAS, 2018) which is executed on the framework Tensorflow. The Tensorflow library is opensource and specific to use in machine learning where a large variety of tasks can be applied (TENSORFLOW, 2018). Another advantage of using the Tensorflow is that it can be run on multiple CPUs and GPUs and it is available for many platforms. From the input values sampled, 70% of them were used to train the classifier, 30% to test the data.

An algorithm was made to calculate the probability for every failure using three classifiers: logistic regression, linear SVM, and multi-layer perceptron. A free software machine learning library *scikit-learn* was used. It is an open-source machine learning library for Python that includes several classifications, regression, and grouping algorithms including supporting vector machines, designed for interaction with Python’s numerical and scientific libraries (SCIKIT-LEARN, 2020).

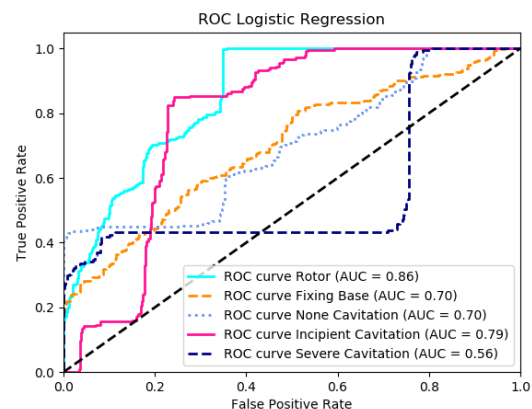
Regarding the implementation in a commercial product, an SoC (System on a Chip) containing, for example, an ARM Cortex A53 processor, is capable of executing the algorithms presented in the article with low power consumption, less than 1 W (NXP, 2020).

It is relevant to highlight that the training of the models is a computationally heavy step, but the execution of the developed models is not. Concerning computational costs, these depend on which processor will be deployed in a commercial product, leaving the processing limitation less relevant than the memory for storage of samples.

Results

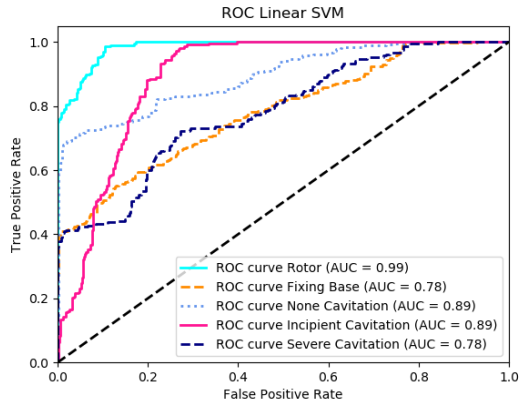
Using the *dataset A*, the logistic regression results are shown in Figure 11. For the linear SVM the results are shown in Figure 12.

Figure 11 – ROC curve for accuracy into fault diagnosis using *logistic regression* as classifier for the *dataset A*.



Source: The authors.

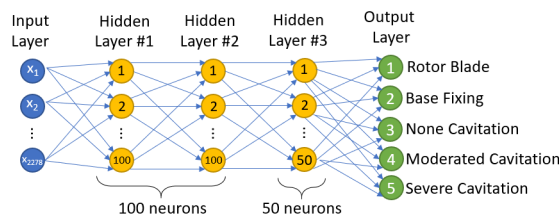
Figure 12 – ROC curve for accuracy into fault diagnosis using *Linear SVM* as classifier for the *dataset A*.



Source: The authors.

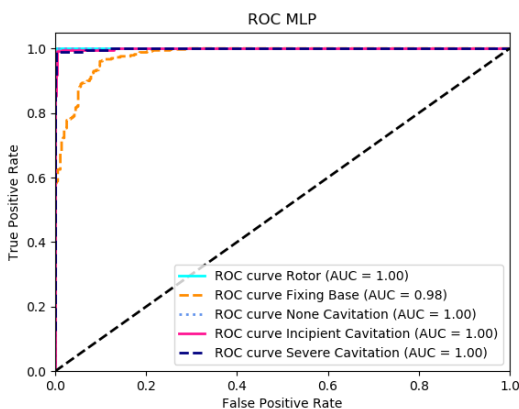
Several neural networks were trained, with a different number of neurons and layers. A network which presented satisfactory results was one with three layers with 100 neurons in the first and second layers and 50 neurons in the last layer, using the same input presented in section *Data Analysis*. The ANN-MLP has five outputs, as shown in Figure 13. Using the ANN-MLP the results are shown in Figure 14.

Figure 13 – ANN-MLP representation with its respective inputs, hidden layers and outputs.



Source: The authors.

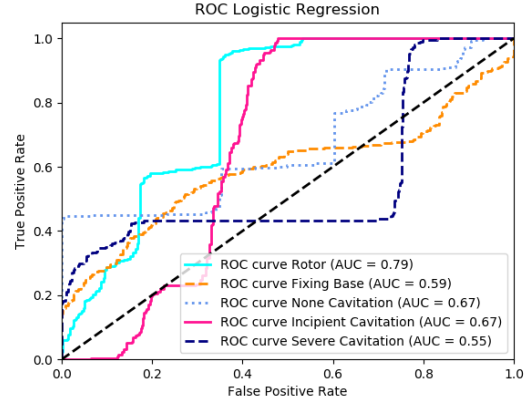
Figure 14 – ROC curve for accuracy into fault diagnosis using *multi-layer perceptron* as classifier for the *dataset A*.



Source: The authors.

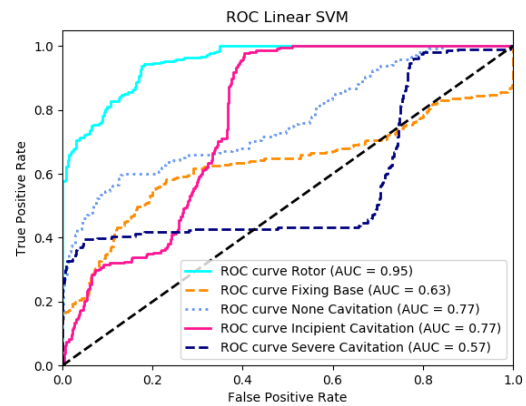
Using the *dataset B*, logistic regression the results are shown in Figure 15, for the linear SVM the results are shown in Figure 16 and for ANN-MLP the results are shown in Figure 17.

Figure 15 – ROC curve for accuracy into fault diagnosis using *logistic regression* as classifier using *dataset B*.



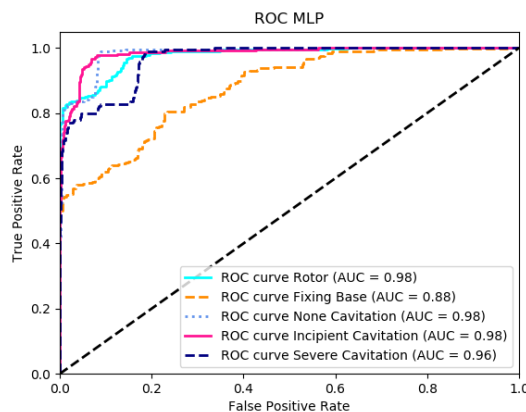
Source: The authors.

Figure 16 – ROC curve for accuracy into fault diagnosis using *Linear SVM* as classifier using *dataset B*.



Source: The authors.

Figure 17 – ROC curve for accuracy into fault diagnosis using *multi-layer perceptron* as classifier using *dataset B*.



Source: The authors.

Comparative results

The Table 4 summarizes the results obtained for each classifier applied to diagnose the centrifugal pump failures for the test data. In table, the result of prediction accuracy using the three classifiers is presented. A comparison with the input type also was made and the difference between

Table 4 – Comparing the results obtained using different number of inputs.

Faults	Prediction Accuracy (%)								
	Logistic Regression			Linear SVM			ANN-MLP		
	Whole Spectrum dataset A	10th firsts Harmonics dataset B	Diff.	Whole Spectrum dataset A	10th firsts Harmonics dataset B	Diff.	Whole Spectrum dataset A	10th firsts Harmonics dataset B	Diff.
Rotor Blade	82.0	79.5	2.5	91.2	85.2	6.0	100.0	89.0	11.0
Fixation Base	63.2	56.0	7.2	69.9	64.1	5.8	92.4	77.8	14.6
None Cavitation	62.3	62.3	0.0	86.8	76.1	10.7	99.7	94.0	5.7
Incipient Cavitation	67.9	68.1	0.1	71.9	69.4	2.5	99.1	92.3	6.8
Severe Cavitation	69.6	69.6	0.0	75.9	69.6	6.3	99.4	88.3	11.1

Source: The authors.

the result obtained for the whole spectrum and the 10th first harmonics was computed. In this case, the datasets did not result in a significant accuracy difference.

It was obtained more than 56% accuracy in predict the probability of each failure using logistic regression and linear SVM. Using the ANN-MLP classifier was possible to achieve an average accuracy of 99.4%. Also, it is observed that the highest difference found in the comparisons was using the MLP, whose highest value was 10.3%. However, it was possible to obtain prediction accuracy value over then 77% using *dataset B* and over then 92% using *dataset A*. In both datasets, the MLP presented better results.

The time spent to execute the ANN-MLP in a personal computer using the *dataset A* as input was 94 s and for the *dataset B* was 67 s. Using the *dataset B* the computational analysis is made about 28% faster. It may be advantageous to use a dataset that provides a reliable response and a shorter processing time. Thus, process actions can be done more quickly to minimize equipment damage.

The large number of harmonics used in *dataset A* brought more information than the low number used in *dataset B*. However, this also increased the computational complexity of the method. Choosing the complexity of the method to be committed is a user choice that depends on the accuracy of the required information versus the existing computational capacity.

Conclusion

This work presented three types of classifiers to analyze vibration in a centrifugal pump using a low-cost MEMS accelerometer. This vibration analysis aimed to

diagnose three types of failures: rotor conditions (new or damage), base fixation (stiff or loose), and the cavitation condition (none, incipient and severe). These diagnoses were made through the following features: calculating FFT from the vibration of the X-axis, vibration RMS values from X, Y, and Z-axis, and the machine speed.

Three classifiers were used: logistic regression, linear SVM, and multi-layer perceptron. The results were compared to evaluate which of them obtained the best performance. It was observed that the ANN-MLP provided better results, being capable of diagnosing with more than 92% accuracy in predicting for the mechanical failures applied, and with about 99% accuracy in evaluating the cavitation status.

The low-cost equipment used to acquire the vibration data was capable to provide enough data to extract features to analyze the centrifugal pump health. Data processing must be done to apply the classifiers. The types of classifiers presented are suggested by the authors. A maintenance manager may use the classifier that is most familiar. However, the results may differ from those presented here.

The feature extraction was considered fundamental to obtain these results. Combine with other possible features, such as temperature and pressure, can improve the predictions could be better yet. However, these variables need invasively install on the test bench.

The objectives were achieved and each of the faults was analyzed. Even the incipient cavitation that is considered hard to identify, could be detected with a high degree of reliability. In future work, vibration measurement can be done at other pump positions simultaneously, bringing another perspective of vibration detection that can lead to even more accurate results.

Acknowledgment

The authors would like to thank the UTFPR campus Curitiba for the training license granted and to DINTER program between UNICAMP / UTFPR Campo Mourão (CAPES 193431/2014).

References

- AUGUSTYN, T. Energy efficiency and savings in pumping Systems: the holistic approach. *Southern African Energy Efficiency Convention*, Johannesburg, p. 1-7, 2012.
- BAZANINI, G. Cavitation erosion pits and craters in metals. *Semina: Ciências Exatas e Tecnológicas*, Londrina, v. 38, n. 2, p. 43, 2018
- BRAUM, S. G. *Mechanical signature analysis: theory and applications*. London: Academic Press, 1986.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, Piscataway, v. 2, n. 4, p. 303-314, Dec. 1989.
- DUTTA, N.; UMASHANKAR, S.; SHANKAR, V. K. A.; *et al.* Centrifugal pump cavitation detection using machine learning algorithm technique. *IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe*, Palermo, p. 1–6, 2018.
- FLORINO, J. A. C.; MELLO, L. F.; CARAZZAI, R. F. Redução de custos operacionais em indústrias de manufatura de MDF. *Semina: Ciências Exatas e Tecnológicas*, Londrina, v. 35, n. 1, p. 77, 2014.
- GOLDBERG, Y.; HIRST, G. *Neural network methods in natural language processing*. Williston: Morgan & Claypool, 2017.
- HAMOND, O.; ALABIED, S.; XU, Y.; DARAZ, A. *et al.*, Vibration based centrifugal pump fault diagnosis based on modulation signal bispectrum analysis. In: CONFERENCE ON AUTOMATION AND COMPUTING (ICAC), 23., Huddersfield, 2017. *Proceedings [...]*. Huddersfield, ICAC, 2017. p. 1-5.
- HARRELL JUNIOR, R. F. E. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. New York: Springer-Verlag, 2001.
- HAYKIN, S. *Neural networks: a comprehensive foundation*. 2. ed. New Jersey: Prentice Hall, 1998.
- HAYKIN, S. *Neural networks and learning machines*. 3. ed. New Jersey: Prentice-Hall, 2008.
- IEEE Guide for Induction Machinery Maintenance Testing and Failure Analysis. *IEEE*, [textitS. l.], p. 1-68, Apr. 2007.
- JARDINE, A. K. S.; LIN, D.; BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, London, v. 20, n. 7, p. 1483-1510, 2006.
- KAYA, D.; YAGMUR, E. A.; YIGIT, K. S.; KILIC, F. C. Kilic *et al.* Energy efficiency in pumps. *Energy Conversion and Management*, Amsterdam, v. 49, n. 6, p. 1662-1673, 2008.
- KERAS. *The Python Deep Learning library*. 2017. Available from: <<https://keras.io/>>. Access in: Aug. 2018.
- KNAPP, R. T.; DAILY, J. W.; HAMMIT, F. G. *Cavitation*. San Francisco: McGraw-Hill, 1970.
- LAROSE, D. T. Logistic regression. In: LAROSE, D. T. *Data Mining Methods and Models*. New Jersey: JOHN WILEY, 2006. p. 155-199.
- MATTOS, E. E.; FALCO, R. *Industrial Pumps*. 2. ed. Rio de Janeiro: Editora Interciência, 1998.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 27, Haifa, 2010. *Proceedings [...]*. Haifa: ICML, 2010.
- NAKAJIMA, S. *Introduction to TPM: total productive maintenance (preventative maintenance series)*. [S. l.]: Eleventh Printing edition, 1988.
- NASIRI, M. R.; MAHJOOB, M. J.; VAHID-ALIZADEH, H. Vibration signature analysis for detecting cavitation in centrifugal pumps using neural networks. In: INTERNATIONAL CONFERENCE ON MECHATRONICS, 2011, Istanbul. *Proceedings [...]*. Istanbul: IEEE, 2011. p. 632-635.
- NORTON, M.; KARCZUB, D. *Fundamentals of noise and vibration analysis for engineers*. 2. ed. Cambridge: University Press, 1994.
- NXP. *MMA8451Q, 3-axis, 14-bit/8-bit digital accelerometer*. Netherlands: NXP, 2018. Available from: <<https://www.nxp.com/docs/en/datasheet/MMA8451Q.pdf>>. Access in: Aug. 2018.

- NXP, *MMA8451Q: $\pm 2g/\pm 4g/\pm 8g$, Low g, 14-bit Digital Accelerometer*. Netherlands: NXP, 2017. Available from: <http://www.nxp.com/products/sensors/accelerometers/3-axis-accelerometers/2g-4g-8g-low-g-14-bit-digital-accelerometer:MMA8451Q?tab=Buy_Parametric_Ta>. Access in: Mar. 2017.
- NXP, *i.MX 8M Quad Power Consumption Measurement*. Netherlands: NXP, 2020. Available from: <<https://www.nxp.com/docs/en/nxp/application-notes/AN12118.pdf>>. Access in: Nov. 2020.
- PEDOTTI, L. A. S.; ZAGO, R. M.; FRUETT, F. Fault diagnostics in rotary machines through spectral vibration analysis using low-cost MEMS devices. *IEEE Instrumentation & Measurement Magazine*, New York, v. 20, n. 6, p. 39-44, 2017.
- PEDOTTI, L. A. S. *Dispositivo IoT de baixo custo para Diagnóstico de Falhas em Máquinas Rotativas*. 2019. Tese (Doutorado) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2019.
- POWERS, D.M.W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, Singapore, v. 2, n. 1, p-37-63, 2011.
- RAO, S. *Mechanical vibrations*. 5. ed. New Jersey: Prentice Hall, 2010.
- ROJAS, R. *Neural networks: a systematic introduction*. New-York: Springer-Verlag, 1996.
- ROJO-ÁLVAREZ, J. L.; MARTINEZ-RAMÓN, M.; MUÑOZ-MARÍ, J.; CAMPS-VALLS, G. Support Vector Machine and Kernel Classification Algorithms. *Digital Signal Processing with Kernel Methods*, [S. l.], p. 433-502, 2018.
- SÁNCHEZ, W.; CARVAJAL, C.; POALACIN, J. *et al.* Detection of cavitation in centrifugal pump for vibration Analysis. INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION AND ROBOTICS (ICCAR), 4., 2018, Auckland. *Proceedings* [...]. Auckland: ICCAR, 2018. p.460-464.
- SCIKIT-LEARN. *An introduction to machine learning with scikit-learn*. 2018. Available from: <<http://scikitlearn.org/stable/tutorial/basic/tutorial.html>>. Access in: Aug. 2018.
- SCIKIT-LEARN developers. *Scikit-learn: Machine Learning in Python*. Available from: <<https://scikit-learn.org/stable/>>. Access in: Nov. 2020.
- STEEGE, P. Overall equipment effectiveness in resist processing equipment. In: ADVANCED SEMICONDUCTOR MANUFACTURING CONFERENCE AND WORKSHOP – ASMC, San Diego, 1996. *Proceedings* [...]. San Diego: IEEE, 1996. p. 76-79.
- STOPA, M. M; CARDOSO FILHO, B. J.; MARTINEZ, C. B. Incipient Detection of Cavitation Phenomenon in Centrifugal Pumps. *IEEE Transactions on Industry Applications*, New York, v. 50, n. 1, p. 120-126, 2014.
- SCHRODER, F; LUCCA, Y.; DALFRÉ FILHO, J. G. Bearing Temperature effect analysis of centrifuge pumps operating with moderate cavitation. *American Journal of Hydropower, Water and Environment Systems*, [S. l.], 2015.
- TENSORFLOW.ORG. *An open source machine learning framework for everyone*. 2018. Available from: <<https://github.com/tensorflow/tensorflow>>. Access in: 4 Nov. 2020.
- WHITE, F. M. *Fluid Mechanics*, 8. ed. Island: University of Rhode Island, 2016.
- WOWK, V. *Machinery vibration: measurement and analysis*. New York, McGraw-Hill, 1991.

Received: Sept. 08, 2020

Accepted: Nov. 04, 2020