Proceeding Paper

Intelligent Fault Diagnosis of Centrifugal Pump Valves in Microbreweries †

Marcio Rafael Buzoli [1,*], **Matheus Luis Despirito** [1], **Fabio Romano Lofrano Dotto** [1], **Pedro de Oliveira Conceição Junior** [1] **and Andre Luis Dias** [2]

1   Department of Electrical and Computer Engineering, São Carlos School of Engineering (EESC), University of São Paulo (USP), São Carlos 13566-590, SP, Brazil; m.l.despirito@usp.br (M.L.D.); fabio.dotto@usp.br (F.R.L.D.); pedro.oliveiracjr@usp.br (P.d.O.C.J.)
2   Electrical and Computing, Federal Institute of São Paulo (IFSP), Sertãozinho 14169-263, SP, Brazil; andre.dias@ifsp.edu.br
*   Correspondence: buzoli@usp.br; Tel.: +55-16-9357-2205
†   Presented at the 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: The brewing industry is expanding with the rise of many small breweries. These are typically small and medium-sized enterprises producing a few hectoliters of beer per batch, often with limited investment capacity for equipment. Centrifugal pumps play a crucial role in microbreweries, facilitating the movement of wort throughout various stages of the brewing process. Failures in these systems, such as valve positioning issues or blockages, can lead to longer production times, increased energy consumption, and potential quality issues. This study explores a soft sensor approach for developing IFDs (Intelligent Fault Detection systems) by using pump drive data—current, torque, and power factor—without the need for additional sensors. Data were collected via a managed switch, and models were trained using Support Vector Machine and Multilayer Perceptron algorithms. The results indicate that this IFD method holds great potential for enhancing automation and maintenance in small breweries.

Keywords: intelligent fault diagnosis; Artificial Neural Network; Support Vector Machine; centrifugal pump; microbreweries

1. Introduction

In recent years, the brewing industry has stood out, particularly with the growth of microbreweries, which have significantly contributed to product diversification and innovation in the sector, thus boosting the brewing landscape in Brazil. The Brazilian brewing market has experienced continuous and significant growth over the last few years. In 2019, the country had 1209 registered breweries, a number that increased to 1847 in 2023, representing a 52.77% increase during this period, demonstrating the resilience of the sector despite the economic challenges posed by the COVID-19 pandemic. This expansion highlights not only the strength of the industry but also the increasing consumer demand for craft beer and innovative products, which has driven both domestic market growth and international interest [1].

Given this scenario, the early detection of failures in centrifugal pumps can help ensure the continuity and efficiency of production processes. Intelligent Fault Detection (IFD) plays a key role in monitoring the operational health of machines and equipment, flagging potential anomalies in processes and allowing for the implementation of corrective measures that prevent greater losses [1]. This method, increasingly popular in condition monitoring, combines physical sensors with software models, using easily measurable variables to estimate process parameters that would otherwise be costly or difficult to measure due to technical limitations, measurement delays, or complex environments [2].

These technological advancements have the potential to expand the use of Intelligent Fault Detection systems in industrial applications, where they are critical for ensuring reliability and operational health in production processes [3]. The aim of this study was to investigate and develop predictive systems capable of efficiently identifying failures related to pump inlet and outlet blockages using machine learning techniques. The challenge lies in the fact that this approach eliminates the need for additional or specialized sensors, making it particularly appealing to microbreweries, which operate with limited resources for advanced automation, and aligns with Industry 4.0 trends.

2. Background

2.1. Intelligent Fault Diagnosis

Intelligent Fault Detection (IFD) refers to the application of machine learning theories to diagnosing machine failures. This method aims to reduce the reliance on human labor by automatically recognizing the health states of machines. Traditionally, fault diagnosis heavily depended on the experience and knowledge of engineers; however, with advancements in machine learning theories such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), it has become possible to develop diagnostic models that learn from collected data. These models can establish a relationship between monitoring data and machine health states, minimizing human intervention [4].

2.2. Support Vector Machine

The Support Vector Machine (SVM) is a widely used supervised learning model for classification, particularly efficient in recognizing health states, such as diagnosing failures in pumps, motors, and other mechanical systems. Initially, the SVM algorithm could only classify two linearly separable classes of data, known as hard-margin SVM. Over time, the algorithm was improved, allowing it to classify non-linearly separable and noisy data, through the introduction of the kernel function (K) concept and the implementation of the box constraint (C) to minimize classification errors [5].

These improvements made SVM more robust in handling noise and outliers by optimizing the hyperplane's position through the adjustment of the C hyperparameter and the Gaussian kernel [4].

$$K(x_{in}, y_{in}) = e^{-\gamma |x_{in}, y_{in}|^2} \tag{1}$$

2.3. Multilayer Perceptron

Artificial Neural Networks (ANNs), inspired by the functioning of the human brain, are a powerful tool for diagnosing machine failures. They consist of multiple layers of interconnected neurons, where input data pass through hidden layers and are mapped to a desired output. Each neuron applies an activation function that processes the received data and transmits the data to the following neurons. During training, the weights of these connections are adjusted to minimize the error between the network's prediction and the actual response. This process allows the ANN to learn to identify complex patterns, making it effective in various diagnostic applications.

3. Proposed Method

3.1. The Microbrewery Pilot Plant

The project was developed in the brewing pilot plant at the Federal Institute of São Paulo, Sertãozinho campus. The control panel was adapted to collect RMS electric current, torque, and power factor data from the centrifugal pump, which were provided by the electric drive and sent to the plant's programmable controller via a PROFINET communication network. The frequency inverter controls the operation of the centrifugal pump according to a rotational setpoint indicated by the operator through the programmable controller, and all of these data are made available as feedback. These data points are often discarded or underutilized but can be applied to detect and diagnose the investigated faults in the system. Figure 1 illustrates the schematic of the adaptation made to the control panel of

the brewing pilot plant. Essentially, the adaptation involves configuring a managed switch with port mirroring capability, enabling it to copy the packets exchanged between the PLC and the frequency inverter. The components of the control panel used in this adaptation are detailed in Table 1.
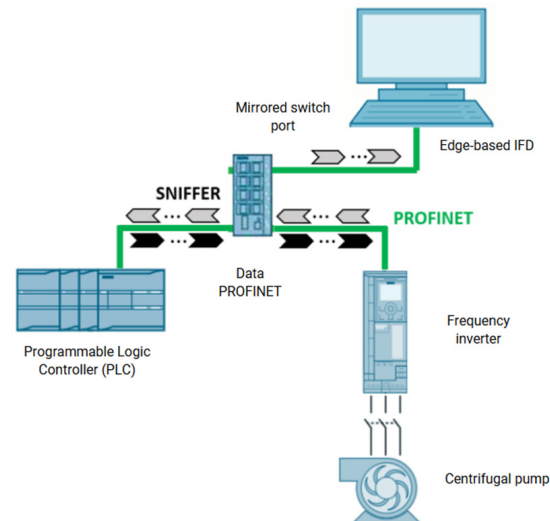

Figure 1. Control panel of the brewing pilot plant.

Table 1. Components of the pilot microbrewery's control panel.

	Item	Feature
1	PLC	PLC S71200 CPU1214C
2	Frequency Inverter	SINAMICS G120
3	Sniffer	SCALANCE TAP104
4	Motor/Pump	Hydrobloc P500
5	Edge-based IFD	Wireshark with Python environment

3.2. Data Collection Description

For data collection, the centrifugal pump from the wort filtration stage was used. During this stage, the pump is either used to recirculate the wort or transfer it to the lautering stage. The pump was operated under a healthy condition and two fault conditions: one where the inlet was blocked and another where the outlet was blocked, both simulated by a valve. Data from the PROFINET network were collected using a sniffer, with the Wireshark software, version 4.2.5, capturing network traffic and saving it in .PCAP files. To interpret these data, an algorithm was developed in Python using the Scapy library to process the .PCAP files. Figure 2a presents an example of raw PROFINET data, while Figure 2b shows the corresponding data in Wireshark, highlighting information such as current speed, electric current, torque, and power factor.

The data from the frequency inverters used in the centrifugal pumps are expressed as percentages relative to the nominal values set in the engineering tools. These values can range from −200% to 200% of the configured nominal value, depending on the parameter being analyzed. It is worth noting that the data interpretation follows the PROFIdrive profile of the PROFINET protocol, although it can be easily adapted to other real-time Ethernet protocols.

To build the dataset, all packets sent within a one-second time interval were considered. Given that the inverter has an update rate of one millisecond, each sample accounts for a thousand data points for each variable investigated. A feature extraction step was applied to these raw signals. For this, a specific algorithm was developed to extract statistical

features from the signals, ultimately enabling the creation of the dataset. The extracted attributes are presented in Table 2.
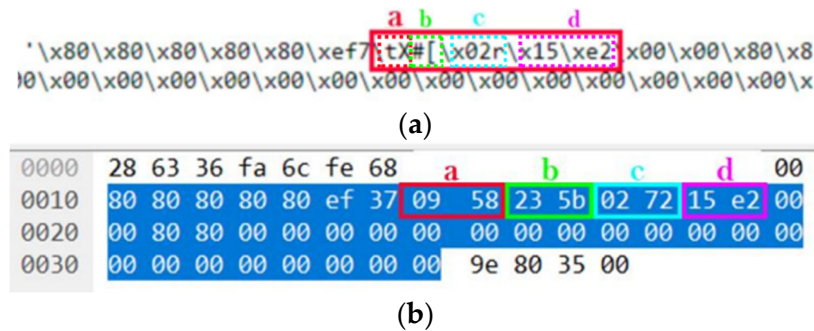

(a)

(b)

Figure 2. Data attributes from a PROFINET data packet: a = Current Speed, b = Electric Current, c = Motor Torque, d = Power Factor.

Table 2. Equations of the attributes investigated for each measured variable.

Feature	Equation	
Mean value	$S_1(x) = \frac{\sum_{i=1}^{n}(x_i)}{n}$	(2)
Standard Deviation	$S_2(x) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - mean\ (x_i))^2}{n-1}}$	(3)
Max. Value	$S_3(x) = max|x_i|$	(4)
Difference between max. and min. value	$S_4(x) = max|x_i| - min|x_i|$	(5)
Entropy	$S_5(x) = \frac{\sum_{i=1}^{n}log_{10}x_i^2}{n-1}$	(6)
Min. Value	$S_6(x) = min|x_i|$	(7)
Kurtosis	$S_7(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x_i))^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x_i))^2\right]^2}$	(8)
Root Mean Square (RMS)	$S_8(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|x_i|}$	(9)
Skewness	$S_9(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \bar{x})^3}{\sigma^3}$	(10)

These attributes were extracted from the RMS electric current, torque, and power factor signals. The current speed value was not considered, as the inverter used does not have an encoder, making it impossible to obtain an accurate rotational speed measurement. Initially, we analyzed the raw data to identify exploratory opportunities, examining 1 s operation samples for the three selected attributes (current, torque, and power factor) in each operation of the experimental setup (normal condition, inlet blockage, and outlet blockage).

Figure 3 illustrates the system's behavior under normal conditions, where the variable values remain within the expected limits, showing a stable and consistent pattern. However, when comparing this condition to Figures 4 and 5, which represent the inlet and outlet blockage failure conditions, respectively, a significant difference in values can be observed. In failure scenarios, the graphs exhibit more pronounced fluctuations, with considerable variations in the monitored parameters, reflecting the negative impact of the blockages.

An interesting observation can be made when comparing the graphs in Figures 4 and 5. Despite representing failures in different parts of the system (inlet and outlet), the patterns in the variable behavior are quite similar. This suggests that both inlet and outlet blockages have comparable effects on the system's performance, resulting in data characteristics that, in many cases, may be visually difficult to distinguish. These graphs highlight how both failure conditions produce similar anomalies in the monitored variables, even though the physical causes of the issues are distinct.
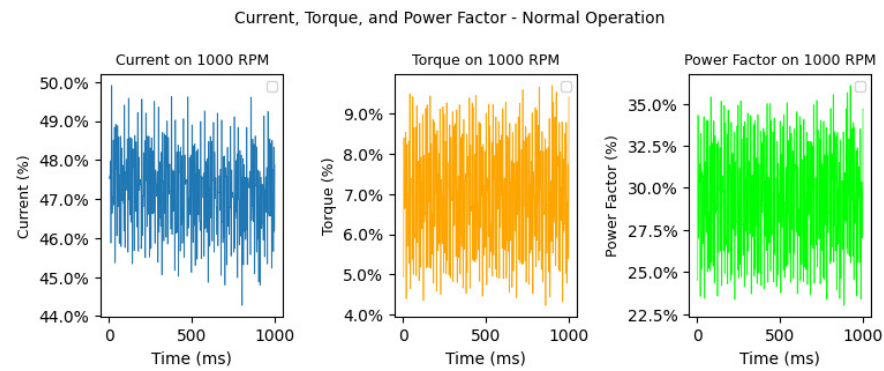
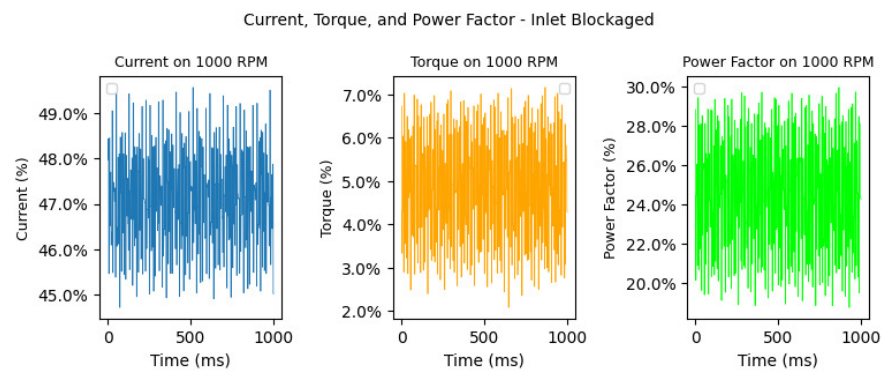**Figure 3.** Behavior of the variables during 1 s of sampling under normal conditions at 1000 RPM.

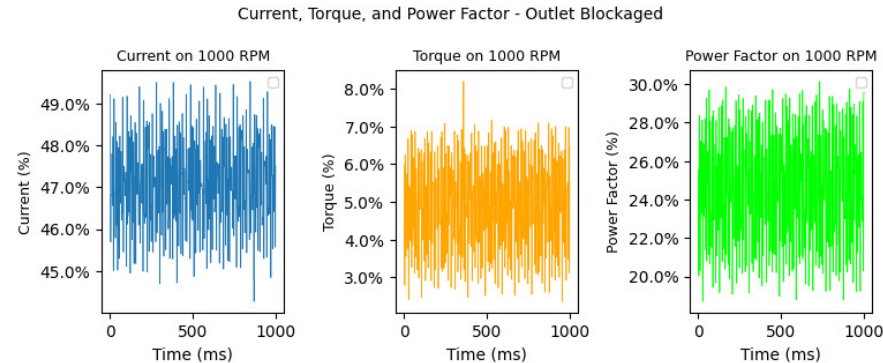**Figure 4.** Behavior of the variables during 1 s of sampling with inlet blocked at 1000 RPM.

Figure 5. Behavior of the variables during 1 s of sampling with outlet blocked at 1000 RPM.

3.3. Training the IFDs

Before training the models, the data were normalized to standardize the variable scales, which is an essential step to avoid features with different magnitudes disproportionately influencing the algorithm. Below, you can observe the PCA (Principal Component Analysis) graph in Figure 6, which works by identifying directions (called principal components) along which the data exhibit the most variation. These directions are linear combinations of the original variables, and the PCA ranks these directions based on how much variance (information) they capture.

The dataset was divided into 80% for model training and 20% for testing, selected randomly. For the SVM-based models, the soft margins functionality was used, with hyperparameters C = [0.01, 0.1, 1, 10, 100] and a Gaussian kernel, where the kernel scale parameter gamma = [0.01, 0.1, 1, 10, 100] was also considered.

For the Multilayer Perceptron models, two hidden layers were used. The training algorithm employed was Adam (Adaptive Moment Estimation), an efficient optimizer that adjusts the learning rate adaptively for each parameter in the network, enabling faster and

more stable convergence. The hidden layers used the ReLU activation function, and the output layer used the sigmoid function, ideal for predicting the two classes. The number of neurons in both hidden layers was set to n = [1, 5, 10, 15, 20]. Performance evaluation for the SVM and ANN (MLP) algorithms was based on key indicators, such as accuracy, false positive rate (FPR), and false negative rate (FNR). Accuracy measures the proportion of correct classifications relative to the total samples, while FPR indicates how often incorrect classifications were made as positive. FNR, on the other hand, reflects the proportion of positive cases that were not detected. These metrics are crucial for understanding how well the models perform in different classification scenarios.
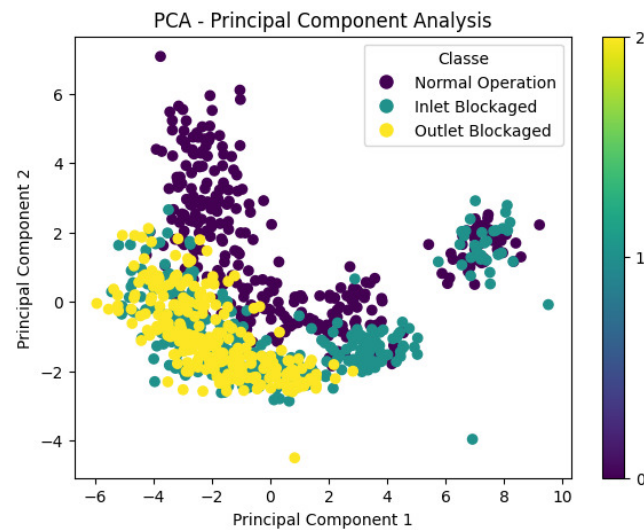

Figure 6. Principal Component Analysis.

3.4. Performance Evaluation

Performance evaluation for the SVM and ANN (MLP) algorithms was based on indicators like accuracy, false positive rate (FPR), and false negative rate (FNR). Accuracy measures the proportion of correct classifications relative to the total samples, while FPR indicates how often incorrect classifications were made as positive. FNR reflects the proportion of positive cases that were not detected.

4. Results and Discussion

Table 3 summarizes the performance metrics, including accuracy, false positive rate (FPR), false negative rate (FNR), the gamma parameter (γ), which represents the kernel scale, and the smoothing coefficient C, which controls the penalty for classification errors. The models were tested using different feature combinations, ranging from all 27 features to specific subsets like current, torque, and power factor. The model using all features performed the best, achieving 90.15% accuracy, with a gamma value of 0.01 and a smoothing coefficient of 10.

Table 3. Results for SVM-based models.

	N. of Features	Feature	Accuracy	FPR	FNR	γ	C
All	27	F1 to F27	90.15%	11.90%	8.40%	0.01	10
Current	9	F1 to F9	75.37%	27.38%	22.69%	0.01	100
	1	F1	75.86%	19.05%	27.03%	100	10
Torque	9	F9 to F18	87.19%	20.24%	7.56%	0.1	100
	1	F10	79.80%	38.10%	7.56%	0. 1	0.1
Power Factor	9	F19 to F27	89.66%	15.48%	6.72%	0.1	100
	1	F19	87.68%	13.10%	11.76%	1	0.1

The results obtained with Artificial Neural Network (ANN) models were evaluated using the same feature sets as SVM. Table 4 summarizes the performance metrics, including accuracy, false positive rate (FPR), false negative rate (FNR), and the number of neurons in the hidden layers, represented by N1 and N2. The model using the full feature set achieved an accuracy of 89.66%, with an FPR of 10.71%, configured with 15 neurons in the first hidden layer and 10 neurons in the second layer.

Table 4. Results for ANN-based models.

	N. of Features	Feature	Accuracy	FPR	FNR	N1	N2
All	27	F1 to F27	89.66%	10.71%	6.28%	15	10
Current	9	F1 to F9	75.86%	22.62%	14.35%	10	1
	1	F0	74.88%	9.52%	19.37%	15	20
Torque	9	F9 to F18	88.18%	16.67%	5.29%	20	20
	1	F10	79.31%	30.95%	8.21%	1	20
Power Factor	9	F19 to F27	90.64%	10.71%	5.29%	15	15
	1	F19	87.68%	13.10	7.25%	1	5

5. Conclusions

The results show that the SVM model with 27 features achieved an accuracy of 90.15%, while the ANN model, with the same attributes, reached an accuracy of 89.66%. Although both demonstrate competitive performances, the SVM slightly outperformed in terms of overall accuracy. However, when isolating the power factor, the ANN model achieved 90.64% accuracy, slightly higher than the SVM, suggesting that this attribute is particularly relevant for the task.

The analysis of a reduced feature space, such as the power factor, consumes fewer computational resources than running models with 27 attributes. The best individual results came from the averages of current (F0), torque (F9), and power factor (F18). In the SVM, tuning parameters such as gamma (γ) and C are critical, just as the neural network structure, including the number of neurons, is essential in ANN models. The possibility of using fewer attributes without significant loss in performance enhances processing efficiency and suggests that simpler models may be preferable in practical applications

Additionally, the implementation of diagnostic techniques, as discussed in this study, can be relevant to the context of Industry 4.0. The use of sniffers for communication data collection can be complemented by IIoT sensors, which enable real-time extraction of operational information from the Profinet network, such as performance metrics and operating conditions of centrifugal pumps. This integration facilitates cloud-based data processing, where advanced analyses can be conducted to predict failures and optimize maintenance. In this way, companies can not only enhance operational efficiency but also implement proactive predictive maintenance strategies, ultimately reducing costs and improving data-driven decision-making.

Future research could benefit from techniques like deep learning and transfer learning, as they allow the extraction of more complex features and the adaptation of pre-trained models to new tasks with minimal data. These approaches could improve the accuracy and robustness of the models, particularly in challenging scenarios with limited datasets.

Author Contributions: Conceptualization, A.L.D.; methodology, M.R.B. and M.L.D.; software, M.R.B. and F.R.L.D.; validation, F.R.L.D., P.d.O.C.J. and A.L.D.; formal analysis, A.L.D.; investigation, M.L.D.; resources, M.L.D.; data curation, M.R.B. and M.L.D.; writing—original draft preparation, M.R.B.; writing, review and editing, M.R.B., P.d.O.C.J. and A.L.D.; visualization, M.L.D., F.R.L.D. and P.d.O.C.J.; supervision, A.L.D.; project administration, A.L.D.; funding acquisition, F.R.L.D. All authors have read and agreed to the published version of the manuscript.

References

1. Park, Y.-J.; Fan, S.-K.S.; Hsu, C.-Y. A Review on Fault Detection and Process Diagnostics in Industrial Processes. *Processes* **2020**, *8*, 1123. [CrossRef]
2. Zhu, X.; Rehman, K.U.; Wang, B.; Shahzad, M. Modern Soft-Sensing Modeling Methods for Fermentation Processes. *Sensors* **2020**, *20*, 1771. [CrossRef] [PubMed]
3. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
4. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
5. Dias, A.L.; Turcato, A.C.; Sestito, G.S.; Rocha, M.S.; Brandão, D.; Nicoletti, R. A New Method for Fault Detection of Rotating Machines in Motion Control Applications Using PROFIdrive Information and Support Vector Machine Classifier. *J. Dyn. Syst. Meas. Control* **2021**, *143*, 041007. [CrossRef]