Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 42 (2020) 516–523

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019)

Predictive model for the degradation state of a hydraulic system with dimensionality reduction

Elena Quatrini*, Francesco Costantino, Cesare Pocci, Massimo Tronci

Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Via Eudossiana, 18, 00184 Rome, Italy

* Corresponding author. Tel.: (+39) 06 4458 5260. *E-mail address:* elena.quatrini@uniroma1.it

Abstract

In recent years, the optimization in the use of resources has a key role in achieving a bigger marginality, reducing the operative costs. Due to the advances in the data science field, even the maintenance context is living important changes. The predictive maintenance and the condition-based maintenance can overcome the classic traditional maintenance methods, like the time-based maintenance or the corrective maintenance, with respect to the first intervention, reducing the costs for unscheduled maintenance, manpower, or loss of production and extending the useful life of the components. Based on these presuppositions, the paper proposes the development of a predictive model for the degradation state of the components of a complex hydraulic system, with some tests and some suggestions about the dimensionality reduction. The system has four known types of breakdown, with different degrees of severity; moreover, a fifth parameter represents whether the cycle has reached stable conditions or not.

Keywords: Complex hydraulic systems; Dimensionality Reduction; Machine Learning; Condition Monitoring; Decision Forests; Neural Network

1. Introduction

Hydraulic systems are very important, due to common industrial applications, for example in manufacturing or engineering machinery [1]. A correct condition monitoring of hydraulic systems can improve productivity, reducing maintenance costs and preventing the system from further deteriorating, considering that reliability and safety are important issues [2]. In fact, accurate condition-based maintenance or predictive maintenance strategy can help with the safety management of a plant. Condition monitoring is the starting point for correct condition-based maintenance or predictive maintenance, that can guarantee reduction of machine downtime and maintenance costs. Due to the high competitive environment today, it is necessary to decrease operating and support costs, and for this reason it is important not only to identify a fault but the failure state, too [3]. Moreover, the advancement in sensor technologies has simplified and accelerated the development of multisensory systems, widely used during processes monitoring [4]. For hydraulic systems, it is not commonplace for human operators to detect faults or to monitor the condition of a component, for example valves [5]. It is important to underline that a single component, that can be for example a pump, can directly affect with its performance the normal work of the entire hydraulic system [6]. The research starts from the previous research [7] conducted on the same dataset used in this paper; the specific dataset is available on the UC Irvine Machine Learning Repository. The aim of this paper is to implement a machine learning system able to predict the degradation level of four specific components of a hydraulic system, that are the cooler, the valve, the pump, and the hydraulic accumulator. The situation is considered as a classification problem; in fact, for every component there are predefined degradation levels, so the machine learning model task is to determine the correct class of degradation. The paper proposes even an approach for a certain dimensionality reduction, suggesting a specific selection criterion from the initial features to the final features.

2351-9789
10.1016/j.promfg.2020.02.039

The remainder of the paper is organized as follows. Section 2 presents the related literature. Section 3 presents the materials and the methods and Section 4 the case study and the dataset. Section 5 presents the feature engineering and in Section 6 is presented the practical experiment. The discussion of the paper results, the conclusions, and future research are summarized in Section 7.

2. Literature review

As explained in section 1. "Introduction", hydraulic systems are deeply investigated, due to their widespread industrial presence. There is a research field about condition monitoring for hydraulic systems [8,9] because a correct condition monitoring is essential for the characterisation of the health state of the system [10].

For what concerns the topic of predictive maintenance, and maintenance in general, of hydraulic systems, there are specific researches for aviation and aerospace industry. In fact, airline flight operation departments need to correctly manage their expensive assets, hydraulic systems' components are subjected to multiple wear conditions and hydraulic system has a significant impact on safety, too [11,12,13,14,15,16,17].

In some paper, the authors have decided to focus on a specific component of the hydraulic system. Giving some examples, one of these components is the valve; in fact, valves can be considered the core control component of the hydraulic systems and they have an important role in numerous engineering applications [1,18,19]. Another important component is the pump; [20,21] are focused on piston pump, because it is the main component of hydraulic power system, [14] considers the pump in the context of the civil aviation, focusing on the failure of the aero-hydraulic pump, [22,23] approach the pump as a key component for better improving the hydraulic system reliability and [24] proposes an experimental method for the determination of remaining useful life of the aviation hydraulic piston pump.

For what concern predictive maintenance and degradation state prediction, it is clear that a correct identification of the current degradation state of industrial components is a fundamental step for the implementation of condition-based and predictive maintenance approaches [25]. The identification and the integration of key process variables for the evaluation of the equipment degradation state can be considered an important starting point to eliminate potential failures, ensure stable equipment operation and improve the mission reliability of manufacturing systems and the quality of products [26]. [27] underlines that the maintenance priorities are focused on the criticality of assets, related to asset degradation conditions. The degradation state of the components is indispensable for a more accurate prediction of the remaining useful life of the monitored components [28].

Researches about predictive maintenance with a focus on the degradation state of the components have been carried out about bearings [29], transport systems [30], systems which are subject to competing and dependent failures due to degradation and traumatic shocks [31], degrading system modelled by a gamma process [32], industrial processes modelled by using hidden Markov model [33].

Talking about the specific dataset used in this paper, there are previous works about it [7,34,35].

In [7] a systematic approach is developed and evaluated for the automated training of condition monitoring systems for complex hydraulic systems, with important suggestions about the dimensionality reduction and the cycle-based approach. [34] debates about condition monitoring, proposing a statistical condition monitoring system, and the determination of typical faults related to the hydraulic system as well as the sensors, too. [35], finally, suggests specific methods for feature selection and feature extraction.

The aim of this paper, due to the relevance of the degradation states' prediction for the predictive maintenance, is to develop and to test a predictive model for the degradation state of critical components of a hydraulic system. The research will start from the analysis of previous works on the same hydraulic system [7,34,35].

3. Materials and methods

The proposed predictive model is developed and tested using a real case study, adopting both dimensionality reduction and machine learning algorithms. The followed approach is based on testing different situations and combinations, finding the best set up for the model; in fact, different feature selection and feature extraction methods have been tested, using them in correlation with different machine learning algorithms. The objective of the research is to find the optimal predictive model for the addressed problems. The case study and the steps of the research will be presented in the following sections. The overall methodological framework can be resumed in the scheme (Figure 1):

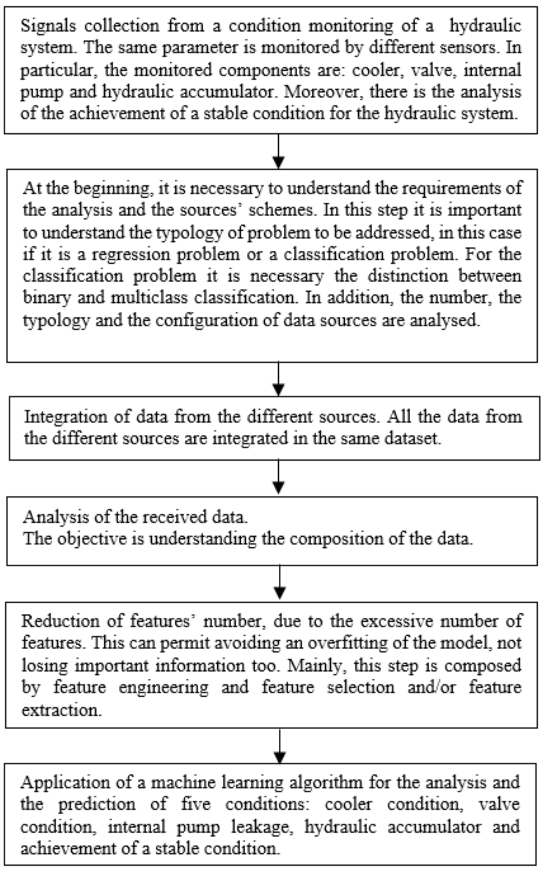

Figure 1. Methodological framework of the research

4.Case study

The core of this paper is an experimental analysis for a complex hydraulic system, that consists of a primary working and a secondary cooling-filtration circuit which are connected via the oil tank.

The features' approach of the paper starts with some suggestions about condition monitoring proposed in [7]. The followed approach proposes both methods for feature selection and feature extraction. The difference between feature selection and feature extraction is that feature selection is for filtering irrelevant or redundant features from your dataset, while feature extraction is for creating a new, smaller set of features that still captures most of the useful information. The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction creates brand new ones. The paper implements, using specific features extracted or selected by the raw data, an analysis of the health state of four components of a hydraulic system and an analysis on the achievement of stable conditions.

The dataset is composed of raw data collected with physical and virtual sensors arranged on four specific components of the hydraulic system: cooler, valve, internal pump, and hydraulic accumulator. They monitor and register some specific process parameters during the load cycles of the hydraulic system. The virtual sensors explain values obtained with physical models based on real value, that are the cooler efficiency, the cooler power, and the system efficiency. The load cycles are 2205 with a duration of 60 seconds for cycle. The paper considers the situation of a fixed working cycle with pre-defined load levels, that is one of the possible working situations [7] because it represents the typical cyclical operations and the load characteristics of industrial applications.

Table 1. Monitored parameters

Physical dimension	Sensor	Units of measure	Frequency
Pressure	PS1, PS2, PS3, PS4, PS5, PS6	Bar	100 Hz
Motor Power	EPS1	W	100 Hz
Flow rate	FS1, FS2	l/min	10 Hz
Temperature	TS1, TS2, TS3, TS4	°C	1 Hz
Vibration	VS1	mm/s	1 Hz
Cooling efficiency	CE	%	1 Hz
Cooling power	CP	kW	1 Hz
System efficiency	SE	%	1 Hz

So, considering every sensor, every monitored parameter and the frequency, the dataset is composed of 43680 features:

- 1 Hz → 8*60 = 480
- 10 Hz → 2*600 = 1200
- 100 Hz → 7*6000 = 42000

As mentioned in the abstract, the system has 4 known types of breakdown. Moreover, for every component there are different levels of performance, from the total efficiency to the breakdown:

- **COOLER CONDITION (%):**
  - 3: Close to total failure (732 cases)
  - 20: Reduced efficiency (732 cases)
  - 100: Full efficiency (741 cases)

- **VALVE CONDITION (%):**
  - 100: Optimal switching behaviour (1125 cases)
  - 90: Small lag (360 cases)
  - 80: Severe lag (360 cases)
  - 73: Close to total failure (360 cases)

- **INTERNAL PUMP LEAKAGE:**
  - 0: No leakage (1221 cases)
  - 1: Weak leakage (492 cases)
  - 2: Severe leakage (492 cases)

- **HYDRAULIC ACCUMULATOR (bar):**
  - 130: Optimal pressure (599 cases)
  - 115: Slightly reduced pressure (399 cases)
  - 100: Severely reduced pressure (399 cases)
  - 90: Close to total failure (808 cases)

In addition to the mentioned breakdowns, the cycle should achieve stable conditions:

- **STABLE FLAG:**
  - 0: Conditions were stable (1449 cases)
  - 1: Static conditions might not have been reached yet (756 cases)

Since the obtainable outputs are categorical, it means that they represent belonging to a specific predetermined category, the presented problems are classification problems. In fact, the classification is the categorisation of the data point in groups.

Cooler condition, valve condition, internal pump leakage, and hydraulic accumulator are multiclass classification problems, because the number of possible outputs is bigger than two; quite the opposite, the stable flag is a binary classification because the possible outputs are two.

5. Feature engineering

This step has been the core step of the analysis; in fact, as explained before, the starting dataset had 43680 features. The excessive number of features involves an important risk of overfitting; the overfitting occurs when a very complex statistical model adapts to the observed data, because it has an excessive number of parameters with respect to the number of observations. The consequence is that the model can guarantee good performance during the training, but not with a new sample of data. Due to the excessive size of parameters, the authors have decided to process the raw data, extracting representative functions for every sensor and every cycle.

The authors have decided, at the beginning, to use six new features, for sensors and cycles, as suggested in [7]. The features are reduced from 43680 to 102 because there are 6 new features for 17 sensors.

The chosen features can be divided into two categories [7]:

- Functions representative of the signal shape:
  - **SLOPE OF LINEAR FIT (SOLF):** it represents, for every sensor and every cycle, the slope obtained by the single measurement for the single sensor.
  - **POSITION OF MAXIMUM VALUE (POM):** for every sensor and every cycle, the authors have computed the maximum value, positioning it in rank with all the other maximum values of the other lines.
- Functions representative of the distribution density characteristics:
  - **MEDIAN (MED):** it represents, for every sensor and every cycle, the value that is assumed by the statistical units in the middle of the distribution.
  - **VARIANCE (VAR):** it represents, for every sensor and every cycle, the variability of the data.
  - **SKEWNESS (SKEW):** it represents, for every sensor and every cycle, the symmetry index of the data.
  - **KURTOSIS (KURT):** it represents, for every sensor and every cycle, a departure from the normal distribution, with respect to which there is a greater flattening or a greater lengthening of the distribution.

So, after all this feature engineering, every line of the dataset represents a cycle, with, for every sensor, the features presented above, as shown in Table 2.

Table 2. Representation of the extracted features

SOLF	POM	MED	SKEW	KURT	…	KURT
PS1	PS1	PS1	PS1	PS1		SE

As explained at the beginning of the paper, the presented problem is a classification problem. In this context, the situation belongs to supervised learning. A supervised learning algorithm learns from labelled training data, and after that, the structured model is able to predict outcomes for unforeseen data. Practically, it means that every line of the dataset needs to be associated with the related output. During the training, the machine learning model learns how to predict the desired outputs only with the input data, that in this paper are presented in Table 2.

The five outputs of the case study are:

- **LABEL 1 (L1): COOLER CONDITION**
- **LABEL 2 (L2): VALVE CONDITION**
- **LABEL 3 (L3): INTERNAL PUMP LEAKAGE**
- **LABEL 4 (L4): HYDRAULIC ACCUMULATOR**
- **LABEL 5 (L5): STABLE FLAG**

The possible values for every output have been presented in section 4. "Case study".

At the end of the labelling phase, the complete dataset for the analysis is (Table 3):

Table 3. Labelled dataset

SOLF	POM	MED	…	L1	L2	L3	L4	L5
PS1	PS1	PS1						

6. Analysis and results

Having processed the original dataset according to [7], the new dataset with its cycle-based features was ready for the implementation. "Cycle-based features" means that the considered features are representative of the situation of a single working cycle. It was decided to use as a development environment a tool offered by the cloud computing service Microsoft Azure, Azure Machine Learning Studio.

Due to the presence of five different classifying labels, the project was divided into five experiments, one for every label.

The experiments came up to be multiclass classification problems all but the last since the fifth label presented only two classes, so it was studied as a binary classification problem, as explained before.

In accordance with [35], it occurred the issue of feature engineering in order to lower the risk of being exposed to overfitting and the 'curse of dimensionality', which would have led respectively to a non-generally applicable solution and a slower computing process.

Following the cited above studies, it was decided to try a set of diverse feature selection and feature extraction methods.

Specifically, these used two main mechanisms: a feature selection method based on the Pearson's correlation coefficient and a feature extraction method known as the Linear Discriminant Analysis (which are going to be referred to as 'Pearson' and 'LDA').

The former was chosen as it is an acknowledged measure of the correlation between two variables, particularly between a feature and the referring label, so the Pearson's method ranks the correlation of every feature with a label, selecting the most correlated ones according to the selection number set.

Moreover, it has been used even in [7] and has a low computational cost, with quick results [35].

In this way, Pearson lowered the dimension of the dataset with just the more relevant features. Pearson's correlation coefficient, for any two variables, represents the value that indicates the strength of the correlation and it is computed by taking the covariance of two variables and dividing by the product of their standard deviations. It is important to mention that the coefficient is not affected by changes of scale in the two variables.

The use of the latter was inspired by previous studies such as [7], which chose LDA for feature extraction; LDA produces in output a new dataset made of linear combinations of the feature variables which can group the data more efficiently into classes, since it is projected in a smaller feature space still preserving the discriminant information. With LDA it is possible to create a new feature dataset that captures the combination of features that best separates two or more classes.

Examples of Pearson's correlation for this case study are present in Figure 2 and Figure 3. An examples for the Linear Discriminant analysis is shown in Figure 4.

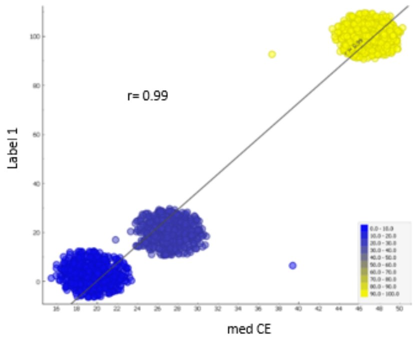

Figure 2. Example of Pearson's correlation for Label 1

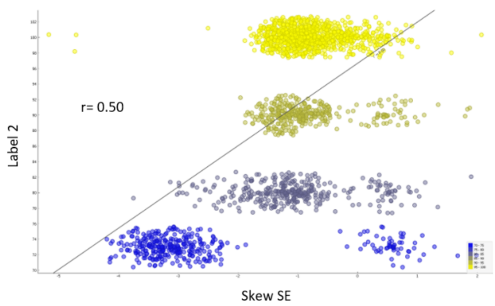

Figure 3. Example of Pearson's correlation for Label 2

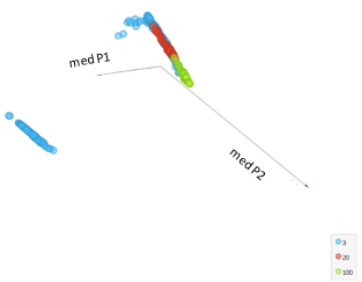

Figure 4. Example of LDA projections for Label 1

Furthermore, these methods were used in four distinct combinations in order to evaluate afterwards the best performing for every experiment.

Two of these processes consisted of a simple application of LDA which was first set to extract 60 features and then 80. The third was a plain utilization of Pearson where the number of features requested to select was 60. In the end, the last one was the only real combination since it employed a sequence of the two methods, selecting and then extracting 60 features.

The number of features received in output from the feature selection and the feature extraction processes (60 or 80) was decided so that it could have been easily replicable and adapted to other datasets with different dimensions. The decision has been structured in that way: the authors have decided to consider for all the known breakdowns and the stable conditions all the features that have a Pearson's correlation with them bigger than 0.15 and the same number of features for the prediction of every breakdown and for the stable conditions and, in this specific case, 60 is the number of features that can guarantee not to neglect features that satisfy this condition. Symmetrically, the authors have decided to use the same number of features with the LDA, testing the difference, only for LDA, increasing the number of features for $\frac{1}{3}$, so at 80. In this way, the size of the final dataset was roughly 60% and 80% respectively, of the one after the first dimensionality reduction. In fact, the first dimensionality reduction creates a dataset with 102 features, the second one with 60 or 80 features respectively. Summarizing, there are two parameters to choose to repeat the proposed methodology: the cut off value for the Pearson's correlation, and the increased percentage of features to test. Further researches will establish optimal values of these two parameters.

The predictive model can be resumed in these steps:

1. Signals collection for the features of interest; at the beginning there are 43680 features
2. Data integration from all the data sources
3. Feature engineering for the first dimensionality reduction; the new features are functions representative of the signal shape or functions representative of the distribution density characteristics. After this step, there are 102 features.
4. Feature selection and feature extraction; in this step, there are two parameters to choose to repeat the proposed methodology: the cut off value for the Pearson's correlation of interest, and the increased percentage of features to test. In this paper, these parameters are:
   - Cut off value = 0.15
   - Increased percentage of features to test = $\frac{1}{3}$

   Different combinations of feature extraction and feature selection methods have been tested:
   - Feature selection with Pearson $\rightarrow$ 60 features
   - Feature extraction with LDA $\rightarrow$ 60 features
   - Feature extraction with LDA $\rightarrow$ 80 features
   - Feature selection with Pearson and then feature extraction with LDA $\rightarrow$ 60 features

   As explained, after this step there are 60 or 80 features.
5. Since the dataset was reduced in dimension, it was subjected to the proper machine learning operation which consisted of two parts: in the first part the algorithms were trained with the 'Train Data', in the second one the algorithms were tested with the 'Test Data', the section of data on which they didn't develop.

In order to get a wider solution, it was established to try four learning models based on the following learning algorithms, setting the models on a different mode according to the nature of the problem, binary or multiclass:

- **NEURAL NETWORK:**
  - Network architecture: fully-connected case

- Number of hidden nodes:100
- The learning rate: 0.1
- Number of learning iterations:100
- The initial learning weights diameter:0.1
- The momentum:0
- The type of normalizer: Min-Max normalizer
- The loss function: CrossEntropy
- The activation function: Sigmoid
- The training algorithm: Back-propagation

- **SUPPORT VECTOR MACHINE:**
  - Number of iterations: 1
  - Lambda: 0.001
  - Normalize features: YES
- **LOGISTIC REGRESSION:**
  - Optimization tolerance: 1E-07
  - L1 regularization weight:1
  - L2 regularization weight:1
  - Memory size for L-BFGS:20
- **DECISION FOREST:**
  - Resampling method: Bagging
  - Number of decision trees:8
  - Maximum depth of the decision trees: 32
  - Number of random splits per node:128
  - Minimum number of samples per leaf node:1

Obviously, these are suggestions for the set up of the algorithms' parameters, that can be set up even in different ways. Further researches will establish optimal values of these parameters.

The usage of the first and second algorithms was based on what was done in [7], so the results got from these resembled those obtained in the older experiment.

The logistic regression algorithm was chosen deliberately to collect a result in output which could have been used as a benchmark for the other outcomes, since it is a widespread algorithm historically acknowledged as a fundamental one in classification problems.

Instead, the selection of a method working with the decision forest algorithm was driven by the will of applying an up-to-date model since its use is becoming more and more prevalent due to its generally satisfying performances. In addition to that, it must be said that this decision was strongly supported by the capability of the algorithm to perform in the presence of noisy features, which could have come up as stated in [34] due to sensor faults or malfunctioning.

In the end, all the models were run on the 'Test Data', thus it was possible to collect a complete set of results, which could project the outcome of each learning algorithm for all the tested conditions in terms of the statistical measure that appeared to be more relevant.

In order to get meaningful results, for the multiclass experiments those were expressed in terms of what was referred to as 'average accuracy', namely the arithmetic mean of every class' accuracy as the ratio between correct predictions and all the predictions; on the other hand, in the binary experiment the chosen metric was the F-Score, that is the harmonic mean of precision and recall, due to the fact that it seemed more pertinent since this label presented an uneven class distribution.

- **F-SCORE:** $2*\frac{Precision*Recall}{Precision+Recall}$

The precision of an algorithm represents the ratio of correct positive observations and the recall is the ratio of correctly predicted positive events.

- **ACCURACY:** $\frac{Number\ of\ correctly\ predicted\ items}{Total\ number\ of\ \ items\ to\ predict}$
- **PRECISION:** $\frac{True\ positives}{True\ positives+\ False\ Positives}$
- **RECALL:** $\frac{True\ Positives}{True\ Positives+False\ Negatives}$

Once the collection of the results was done, the whole set was evaluated, particularly looking for the best combination of feature engineering, feature selection or extraction method and machine learning algorithm in every label, it means the predictive model that can guarantee the best performance for every prediction.

In the first experiment the best outcome was obtained from the application of the method based on the decision forest algorithm on a dataset that was elaborated with Pearson process, although all the results were spread among a few percentage points.

As well as for the first, the classification on the second label got its best score operating with the sequence of Pearson and decision forest, however it is worth to highlight that in this condition the predictive model classified correctly every instance.

In the third experiment the best result was unique, since it was the only one achieved with the neural network algorithm, scoring exactly the same with processes which involved LDA, yet it must be said that almost all the other outcomes were just a few percentage points lower than those mentioned above.

With regards to the fourth label, the learning model which performed better was the one applying the decision forest, similarly to the first couple of experiments, even if it was operated on a dataset that underwent the LDA process set on extracting 80 variables.

In the last one, even if it was a binary classification problem, it was found that the best sequence working on this label was identical to the one in the last described experiment.

In addition to this analysis, it could be stated that combining the feature selection process based only on Pearson and the learning models using support vector machine and logistic regression led to the worst results in every classification, with an exception for the first experiment, in which all the algorithms and all the methods for feature selection and feature extraction led to comparable results.

The results for every tested combination of feature selection or extraction and algorithm are shown in Table 4. An explanation for the columns' name and column S&E is necessary:

- **S&E**→ Method of feature selection or extraction
- **LR** → Logistic Regression
- **NN** → Neural Network
- **DF** → Decision Forest
- **SVM** → Support Vector Machine
- **F60/F80** → LDA with 60/80 features
- **PF60** → LDA and Pearson with 60 variables
- **P60** → Pearson with 60 variables

As explained before, for label from 1 to 4 the performance evaluation is based on the accuracy of the model and for label 5 the evaluation is based on the F-Score.

Table 4. Results

	S&E	LR	NN	DF	SVM
L 1	F80	0.997981	0.997981	0.996971	0.996971
	F60	0.997981	0.997981	0.99596	0.996971
	PF60	0.996971	0.996971	0.997983	0.996971
	P60	0.997981	0.997981	**0.998991**	0.989914
L 2	F80	0.987029	0.999244	0.996967	0.973137
	F60	0.987029	0.998487	0.997726	0.984701
	PF60	0.988533	0.999244	0.996578	0.98549
	P60	0.8556	0.981905	**1**	0.818676
L 3	F80	0.996462	**0.998991**	0.996971	0.995452
	F60	0.996462	**0.998991**	0.99596	0.988849
	PF60	0.998485	**0.998991**	0.997474	0.998485
	P60	0.863574	0.975758	0.995966	0.751601
L 4	F80	0.962462	0.977799	**0.991633**	0.96097
	F60	0.958785	0.968814	0.990859	0.954447
	PF60	0.942539	0.977756	0.987763	0.950126
	P60	0.750443	0.84234	0.986167	0.702836
L 5	F80	0.938326	0.909931	**0.956332**	0.943478
	F60	0.938326	0.907407	0.942478	0.936819
	PF60	0.933045	0.922018	0.945055	0.92569
	P60	0.836601	0.791878	0.935982	0.775599

7. Conclusions

At the end of the paper, the authors can make some reflections. Firstly, the suggested methods about the dimensionality reduction, that are to use Pearson's correlation according to the minimum correlation and the Linear Discriminant Analysis similarly, and the predictive model in general guarantee an important performance. In fact, the suggestion to reduce the number of features at 60 or 80, finding the optimal combination of the inputs that linearly separates each group while minimizing the distances within each group or selecting the *n* variables more correlated with the target label, and the predictive model can guarantee a performance bigger than 99% for the four multiclass classifications and bigger than 95% for the binary classification. Logistic regression and support vector machine with Pearson for label 2, label 3, and label 4 have the worst performance, but in both cases bigger than 80% for label 2, bigger than 75% for label 3 and bigger than 70% for label 4. For label 5 the worst performance is with neural network and support vector machine with Pearson, but in both cases bigger than 77%. Moreover, the authors have tested not only previously used and confirmed algorithms in the same context, it means with the same dataset, but even new algorithms like decision forest and logistic regression. Decision forest is the best algorithm in four situations and only for label 3 neural network outperforms decision forest. There is not a single number of features or a single method of selection or extraction of features with the best performance. In fact, for label 1 and label 2 Pearson with 60 features achieves the best results. For label 3 there is an equal accuracy between LDA with 80 features, LDA with 60 features and the combination between LDA and Pearson with 60 features. For label 4 and label 5 the best performance is achieved by LDA with 80 features. In general, there is a small difference between LDA with 60 features and LDA with 80 features. In fact, both for the multiclass classifications and the binary classification the difference of accuracy and F-Score between LDA with 60 features and LDA with 80 features is never bigger than 0.01 approximatively. Talking about the limitation of the model, that for the authors become future researches, all the selected or extracted features are based on a single cycle. This is a limitation, as mentioned in [7], because there is the risk of neglecting deteriorations visible latter than a cycle. Future research will be focused on adding new features, based on functions representative of the distribution density, computed on a bigger time window. Another future research will be the implementation of the same model in different situations, such as different machines or components. Moreover, on the same hydraulic system, could be useful to try to increase the number of the performance levels or to try different approaches for feature selection or feature extraction.

References

[1] Y. Lei, W. Jiang, A. Jiang, Y. Zhu, H. Niu, and S. Zhang, "Fault Diagnosis Method for Hydraulic Directional Valves Integrating PCA and XGBoost," *Processes*, vol. 7, no. 9, p. 589, 2019.

[2] A. Y. Goharrizi and N. Sepehri, "Application of fast fourier and wavelet transforms towards actuator leakage diagnosis: A comparative study," *Int. J. Fluid Power*, vol. 14, no. 2, pp. 39–51, 2013.

[3] F. Camci and R. B. Chinnam, "Health-state estimation and prognostics in machining processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 3, pp. 581–597, 2010.

[4] K. Liu, N. Z. Gebraeel, and J. Shi, "A Data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, 2013.

[5] A. Steinboeck, W. Kemmetmüller, C. Lassl, and A. Kugi, "Model-based condition monitoring of an electro-hydraulic valve," *J. Dyn. Syst. Meas. Control. Trans. ASME*, vol. 135, no. 6, 2013.

[6] H. L. Tian, H. R. Li, and B. H. Xu, "Fault prediction based on data fusion," in *Advanced Material Research*, 2013, vol. 712–715, pp. 2084–2088.

[7] N. Helwig, E. Pignanelli, and A. Schutze, "Condition monitoring of a complex hydraulic system using multivariate statistics," *Conf. Rec. - IEEE Instrum. Meas. Technol. Conf.*, vol. 2015-July, pp. 210–215, 2015.

[8] J. Nurmi, J. Honkakorpi, J. Vihonen, and J. Mattila, "Micro-electromechanical system sensors in unscented Kalman filter-based condition monitoring of hydraulic systems," in *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics: Mechatronics for Human Wellbeing, AIM 2013*, 2013, pp. 354–361.

[9] B. Chidambaram, D. D. Gilbertson, and K. Keller,

"Condition-based monitoring of an electro-hydraulic system using open software architectures," *IEEE Aerosp. Conf. Proc.*, vol. 2005, pp. 3532–3539, 2005.
[10] S. S. Chawathe, "Condition monitoring of hydraulic systems by classifying sensor data streams," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019*, 2019, pp. 898–904.
[11] W. O. L. Vianna and T. Yoneyama, "Predictive Maintenance Optimization for Aircraft Redundant Systems Subjected to Multiple Wear Profiles," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1170–1181, 2018.
[12] J. Du, S. Wang, L. Han, S. Zhao, and C. Guo, "Prognostic management verification system of aircraft hydraulic power supply system," *IEEE Int. Conf. Ind. Informatics*, pp. 693–699, 2012.
[13] J. P. P. Gomes, L. R. Rodrigues, B. P. Leao, R. K. H. Galvão, and T. Yoneyama, "Using Degradation Messages to Predict Hydraulic System Failures in a Commercial Aircraft," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 1, pp. 214–224, 2018.
[14] G. Xu, C. Ma, Z. Gao, X. Hu, and Y. Luo, "Modeling and simulation of aero-hydraulic pump wear failure," in *2017 Prognostics and System Health Management Conference, PHM-Harbin 2017 - Proceedings*, 2017.
[15] D. Soudbakhsh and A. M. Annaswamy, "Health monitoring with matrix regressor based adaptive observers," *2017 IEEE 56th Annu. Conf. Decis. Control. CDC 2017*, vol. 2018-Janua, no. Cdc, pp. 5131–5136, 2018.
[16] R. Patriarca, F. Costantino, G. Di Gravio, and M. Tronci, "Inventory optimization for a customer airline in a Performance Based Contract," *J. Air Transp. Manag.*, vol. 57, 2016.
[17] F. Costantino, G. Di Gravio, R. Patriarca, and L. Petrella, "Spare parts management for irregular demand items," *Omega (United Kingdom)*, 2017.
[18] W. O. Loesch Vianna, L. G. De Souza Ribeiro, and T. Yoneyama, "Electro hydraulic servovalve health monitoring using fading extended Kalman filter," in *Proceedings of the 2015 IEEE Conference on Prognostics and Health Management (PHM)*, 2015, pp. 1–6.
[19] V. Karanović, M. Jocanović, S. Baloš, D. Knežević, and I. Mačužić, "Impact of contaminated fluid on the working performances of hydraulic directional control valves," *J. Mech. Eng.*, vol. 65, no. 3, pp. 139–147, 2019.
[20] Z. He, S. Wang, K. Wang, and K. Li, "Prognostic analysis based on hybrid prediction method for axial piston pump," *IEEE Int. Conf. Ind. Informatics*, pp. 688–692, 2012.
[21] P. Casoli, M. Pastori, F. Scolari, and M. Rundo, "A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps," *Energies*, vol. 12, no. 5, 2019.
[22] J. Sun, H. Li, and B. Xu, "Prognostic for hydraulic pump based upon DCT-composite spectrum and the modified echo state network," *Springerplus*, vol. 5, no. 1, 2016.
[23] J. G. M. Lázaro, C. B. Pinilla, and S. R. Prada, "A survey of approaches for fault diagnosis in axial piston pumps," in *ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)*, 2016, vol. 4A-2016.
[24] X. Wang, S. Lin, and S. Wang, "Remaining useful life prediction model based on contaminant sensitivity for aviation hydraulic piston pump," in *AUS 2016 - 2016 IEEE/CSAA International Conference on Aircraft Utility Systems*, 2016, pp. 266–272.
[25] P. Baraldi, G. Bonfanti, and E. Zio, "Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 102, pp. 382–400, 2018.
[26] Y. He, C. Gu, Z. Chen, and X. Han, "Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis," *Int. J. Prod. Res.*, vol. 55, no. 19, pp. 5841–5862, 2017.
[27] A. Consilvio, A. Di Febbraro, and N. Sacco, "Stochastic scheduling approach for predictive risk-based railway maintenance," in *IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, 2016, pp. 197–203.
[28] K. A. Kaiser and N. Z. Gebraeel, "Predictive maintenance management using sensor-based degradation models," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 39, no. 4, pp. 840–849, 2009.
[29] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Comput. Ind.*, vol. 106, pp. 14–29, 2019.
[30] J. Foulliaron, L. Bouillaut, P. Aknin, and A. Baros, "A specific dynamic Bayesian network for a prognosis based maintenance strategy," *Civil-Comp Proc.*, vol. 104, 2014.
[31] I. T. Castro, K. T. Huynh, A. Barros, and C. Bérenguer, "A predictive maintenance strategy based on mean residual life for systems subject to competing failures due to degradation and shocks," in *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability*, 2012, vol. 1, pp. 375–384.
[32] E. A. Elsayed and H. Zhang, "Optimum threshold level of degraded structures based on sensors data," in *2006 Proceedings - 12th ISSAT International Conference on Reliability and Quality in Design*, 2006, pp. 187–191.
[33] B. Roblès, M. Avila, F. Duculty, P. Vrignat, S. Bégot, and F. Kratz, "Hidden Markov model framework for industrial maintenance activities," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 228, no. 3, pp. 230–242, 2014.
[34] N. Helwig, E. Pignanelli, and A. Schütze, "Detecting and Compensating Sensor Faults in a Hydraulic Condition Monitoring System," in *SENSOR 2015 - 17th International Conference on Sensors and Measurement Technology*, 2015, pp. 641–646.
[35] T. Schneider, N. Helwig, and A. Schütze, "Automatic feature extraction and selection for classification of cyclical time series data," *Tech. Mess.*, vol. 84, no. 3, pp. 198–206, 2017.