Ensemble of Unsupervised Fuzzy C-Means classifiers for clustering health status of oil sand pumps

Francesco Di Maio, Enrico Zio, M. Pecht, P. Tse, K.L. Tsui

Ensemble of Unsupervised Fuzzy C-Means classifiers for clustering health status of oil sand pumps

F. Di Maio
Energy Department, Polytechnic of Milan, Via Ponzio 34/3, 20133 Milano, Italy

E. Zio
Energy Department, Polytechnic of Milan, Via Ponzio 34/3, 20133 Milano, Italy
Ecole Centrale de Paris and Supelec, Grande Voie des Vignes, 92295 Chatenay-Malabry Cedex, France

M. Pecht, P. Tse & K. Tsui
Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong

ABSTRACT: Detection of anomalies and faults in slurry pumps is an important task with implications for their safe, economical, and efficient operation. Wear, caused by abrasive and erosive solid particles, is one of the main causes of failure. Condition monitoring and on-line assessment of the wear status of wetted components in slurry pumps are expected to improve maintenance management and generate significant cost savings for pump operators. In this context, the objective of the present work is to present a framework for the assessment and measurement of the wear status of slurry pumps when available data is extremely limited. Four sequential steps are performed: data collection, feature extraction, feature selection and classification. The main idea is to combine the predictions of multiple unsupervised classifiers fed with different inputs taken from different signals, based on fuzzy C-means clustering, to reduce the variance of the results so that they are less dependent on the specifics of a single classifier. This will also reduce the variance of the bias, because a combination of multiple classifiers may learn a more expressive concept class than a single classifier.

1 INTRODUCTION

Centrifugal slurry pumps are widely used in the oil sand industry, mining, ore processing, waste treatment, cement production, and other industries to move mixtures of solids and liquids. Equipment manufacturers and owners invest significant resources in maintenance programs designed to ensure that the required hydraulic system performance is maintained at maximum efficiency. In fact, unscheduled outages, costly component replacements and repairs that result from unexpected premature failures or gradual performance degradation caused by system wear can cost companies millions of dollars each year (Mitchell, 1999; Hancock et al., 2006).

The motivation for this application comes from the interest of a producer of oil extracted from oil sands in developing a proper monitoring scheme to detect pump failures in a system aimed at moving large amounts of raw oil sand. The reason for the failures is not exactly known, although it has been conjectured that the main contribution to pump wear is the large flow of oil into the vanes and the presence of small particles of dirt and sand in the sucked fluid (LaBour, 1995; Frith et al., 1996). Previous maintenance and condition monitoring schemes provided insufficient warning of the impending failure. A system that could distinguish between normal machine operation and an impending mechanical failure was needed, i.e., a fault classifier had to be put in operation.

In general terms, fault classification methods can be divided into two categories (Venkatasubramanian et al., 2003): model-based and pattern recognition techniques. In model-based methods, faults can be detected by performing some mathematical calculations. For example, in the case of interest here, the state-of-the-practice entails oil pump failures being diagnosed by expert analysis of the parameter values measured during the monitoring time and their comparison with the nominal power curve of every oil pump: drawing the actual power curve according to the measured parameters values, i.e., by manual analysis, allows the analyst to identify whether any fault exists. Indeed, failed pumps often show hollow pumping action and energy waste. Because of the nonlinearity of the wear behavior and the size of the input data and their uncertainties, this way of proceeding requires significant human, material, and financial resources while not guaranteeing the timely detection of faults, thus seriously affecting production (Tian et al., 2007).

On the other hand, pattern recognition methods offer a framework that can satisfy a number of basic requirements, such as short calculation time, high accuracy, and capability of dealing with nonlinear wear behaviors (Zio, 2007). Especially, soft computing approaches (e.g., Artificial Neural Networks and Fuzzy Logic systems) have shown superior robustness, speed, and accuracy compared to model-based methods (Shahrtash et al., 2008). In pattern recogni-

tion techniques, the conceptual basis for the detection of failure onset is that different system faults initiate different patterns of evolution of the interested variables, as measured by properly placed sensors (Zio et al., 2006).
Pattern recognition methods entail three different stages: feature extraction, feature selection, and classification (Sheng et al., 2004). Figure 1 shows the flowchart of pattern recognition methods: the first step entails the collection into a dataset of raw data, e.g., vibration data; feature extraction consists in the evaluation of the most common summary statistics, e.g. mean, standard deviation, in order to summarize the characteristics of the available data; the aim of feature selection is then to obtain the features which are essential for class separation which is the goal of the last step, i.e., classification.

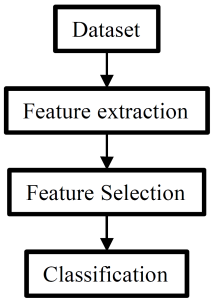

Figure 1. Pattern recognition flowchart

In literature, a number of pattern recognition methods have been proposed that differ in the classification stage, e.g., hierarchical trees (H-trees) (Breiman et al., 1984; Ripley, 1996; Loh et al., 1997), artificial neural networks (ANNs) (Rumelhart et al., 1986; Ripley, 1996; Zhang, 2000), and fuzzy logic (FL) systems (Zadeh, 1965; Klir et al., 1995; Wang et al., 2006; Zio et al., 2006; Wang et al., 2007). All these techniques have been applied to real classification problems in a supervised scheme that entails the classifier to be first trained on data from known faults and then used to classify new data.
H-trees evaluate the contribution of input features in determining the output classes of similarity. Generally, the most effective feature is selected as the first node of the tree, and its border value is used to create two different branches. Then, by the same criterion, the next most effective feature is found in each branch. This process is continued until the final nodes (leaves of the tree) obtained in all of the branches contain only the output classes. Different procedures may be applied to search for the best tree to solve a given problem, and then the best one can be selected by comparing the accuracy of the results and the time required to create the tree (Loh et al., 1997).
ANNs can learn to perform the mapping of the input-output relationships underpinning system behavior by a process of training on many different examples of input and corresponding output states (Rumelhart et al., 1986; Hancock et al., 2006;). A main limitation of ANNs is that the results they deliver are difficult to interpret physically, and thus the underlying model remains cryptic.
FL modeling is designed to handle imprecise linguistic concepts, such as “small”, “big”, “young”, and “low”, and deal with uncertainties (Zadeh, 1965; Zio et al., 2006). FL exhibits an inherent flexibility and has proven to be a successful modeling framework in a variety of industrial applications and pattern recognition tasks (Wang et al., 2006; Wang et al., 2007). One of the main strengths of fuzzy logic modeling compared with other schemes is its capability of dealing with imprecise data (Marseguerra et al., 2004). As for the limitations of fuzzy logic, the main difficulties stand in the fuzzy partitioning of the input and output spaces and in the establishment of the fuzzy rules that are at the basis of the classification phase and may require a time-consuming, trial-and-error process.
In this work, we present a framework for the assessment and measurement of the wear status of slurry pumps when available data is extremely limited. In particular, great efforts are devoted to the design of the classification strategy. In fact, in the case here of interest, the application of supervised classification schemes were precluded due to the unavailability of a comprehensive database of failures. Thus, as we shall see, the approach adopted for detection do not require training. In other words, the classifier is implemented for fault detection in an unsupervised manner, where the training and test phases collapse into the same clustering phase, and the class assignment is automated from available data of unknown classes. In particular, the adopted unsupervised FL approach, i.e., fuzzy clustering, exploits the advantages of automated generation of fuzzy rules, low computational burden, and benefits from the high-level, human-like rule representation typical of fuzzy systems, which offer an appealingly powerful framework for tackling practical classification problems.
Moreover, because of the shortage of data, the robustness of the classification approach is augmented by combining multiple classifiers so as to improve upon the performance of individual classifiers. The idea is to combine the predictions of multiple classifiers (for more details on the methodology, refer to Section V) to reduce the variance of the results and the bias.
The paper is organized as follows. Section 2 presents the case study and the structure of the available database. Section 3 presents the feature extraction step of the pattern recognition process. Section 4 shows the results of box plot analysis for feature selection. The method with which the fuzzy rules are generated from the data set is shown in Section 5. Section 6 reports the results of the method for the classification of the oil pump into failed or safe status, based on

the available vibration data. The monitoring scheme is expected to provide advanced warning and lead time to prepare the appropriate corrective actions. Finally, advantages and limitations of the proposed methodology are discussed.

2 THE CASE STUDY

In this research, experimental data were collected from a number of slurry pumps that are used to deliver a mixture of bitumen, sand, and small pieces of rock from one site to another. For each pump, vibration is monitored as a symptom of system health. Vibration signals have been collected at the inlet and outlet of slurry pumps operating in an oil sand mine. The pump vibration data were collected by the mine staff and one of the authors using the Smart Asset Management System (SAMS) and then further analyzed using the proposed classification methods. SAMS is a PC-based virtual instrument used to perform machine health monitoring (Tse, 2002). Its measurement platform provides a Graphical User Interface that allows the user to choose from different diagnostic techniques (e.g., higher order statistical analysis and orbit analysis) to conduct machinery fault diagnosis. It can be installed in a notebook PC or desktop computer for portable or continuous machine health monitoring. SAMS also provides an easy-to-use interface for data management, report generation, trend analysis, etc., to help the maintenance staff in the recording and planning of maintenance activities.

2.1 *The Database*

The data acquisition equipment (DAQ) consists of a National Instrument (NI) cDAQ 9172 and a DAQ module NI 9234. Their specifications are listed in Table I. Vibrations were measured by four accelerometers mounted in four different positions so that there were a total of four different vibration signals captured. They are denoted as S1, S2, S3, and S4. Accelerometers S1 and S2 were PCB 352A60 accelerometers (see Table I) that were mounted on the case of the pump and denoted as ‘Casing Lower’ and ‘Casing Discharge’, respectively. Accelerometers S3 and S4 were PCB 352C18 accelerometers (see Table I) mounted on the suction and discharge pipes, respectively. All four accelerometers captured the vibration signals from four different positions at a similar sampling frequency rate of 50 kHz.

In Figure 2, the layout of the oil extraction site is represented. It consists of two parallel lines, L1 and L2, each composed of four different pumps. The pumps located in L1 are called G1, G2, G3, and G5, whereas G1, G2, G3, and G4 are those located in L2. Each pump is different in type, size, and working condition, i.e., ground elevation, process fluid, history, and wear.

Table 1. Measurement equipment.

	Equipment	Model	Specification
1	DAQ Device	NI cDAQ 9172	Max. support module = 8
2	DAQ Module	NI 9234	Resolution = 24 bit Input range = +/- 5V Sampling rate = up to 51.2 kHz per channel
3	Smart Asset Maintenance System (SAMS)	Version 2.3.8	
4	Notebook Computer	IBM T60	Intel Core2 processor 1.66 GHz Windows XP Professional
5	Accelerometer	PCB 352A60	Mounted on positions Casing Lower and Casing Discharge Sensitivity = 10.2 mV/g
6	Accelerometer	PCB 352C18	Mounted on positions Suction Pipe and Discharge Pipe Sensitivity = 9.7 mV/g

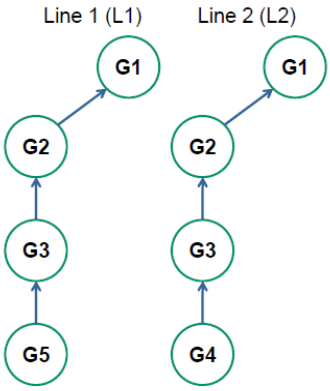

Figure 2. The oil extraction lines layout

Only 11 batches of 4 sensory vibration signals are available in total. The number of patterns for each pump is listed in Table II. These degradation patterns are representative of different stages of progressive pump deterioration. Despite that, in order to analyze only pumps subjected to similar working conditions, we only selected the degradation patterns relative to G1 and G2 from lines L1 and L2. Finally, the total number of available degradation patterns to be classified is 7.

Table 2. List of Available Degradation Patterns for Each Pump in L1 and L2. NA=Not Available.

L1	Available degradation patterns	L2	Available degradation patterns
G1	1	G1	2
G2	1	G2	3
G3	2	G3	NA
G5	1	G4	1

Each degradation pattern is composed of 30 intervals of records, each one lasting 1.3 [s], with pauses of 2 [s] in between (Figure 3).

A preliminary analysis of the data showed that smooth and gradual degradation of the pump performance occurred (except for catastrophic failures), such that there was no significant deviation of the signal along the total 40 seconds of records. Thus, to lighten the computational burden of the data treat-

ment, we have concentrated our analysis only on the records from the first 1.3 [s] (65000 points), discarding the remaining 29 intervals, assuming that the pump is either failed or healthy at time 0 [s].

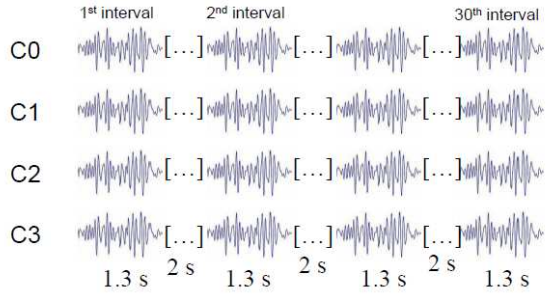

Figure 3. Sketch of the degradation pattern structure.

3 FEATURE EXTRACTION

Ten features were selected and extracted from the batches of vibration data collected by the accelerometers. For each of the 7 degradation patterns, the following M=10 indexes (Lei et al., 2009) were evaluated (N is equal to 65000 sampling points):

(1) Peak value: $\max = \max_{j=1,\ldots,N} n_j$

(2) Mean: $u = \frac{1}{N}\sum_{j=1}^{N} n_j$

(3) Standard deviation: $\sigma = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(n_j - u)^2}$

(4) Root mean square: $RMS = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(n_j)^2}$

(5) Skewness: $SK = \left(\sum_{j=1}^{N}(n_j - u)^3\right) \Big/ \left((N-1)\sigma^3\right)$

(6) Kurtosis: $KU = \left(\sum_{j=1}^{N}(n_j - u)^4\right) \Big/ \left((N-1)\sigma^4\right)$

(7) Crest indicator: $CI = \left(\max_{j=1,\ldots,N}|n_j|\right) \Big/ \sqrt{\frac{1}{N}\sum_{j=1}^{N}(n_j)^2}$

(8) Clearance indicator: $CLI = \left(\max_{j=1,\ldots,N}|n_j|\right) \Big/ \left(\frac{1}{N}\sum_{j=1}^{N}\sqrt{|n_j|}\right)^2$

(9) Shape indicator: $SI = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(n_j)^2} \Big/ \left(\frac{1}{N}\sum_{j=1}^{N}|n_j|\right)$

(10) Impulse indicator: $MI = \left(\max|n|\right) \Big/ \left(\frac{1}{N}\sum_{j=1}^{N}|n_j|\right)$

4 FEATURE SELECTION

The objective of feature selection is three-fold: to improve the performance of the classifier, provide faster and more cost-effective classification, and provide a better understanding of the underlying process that generated the data (Guyon, 2003). Depending on the nature of the regression technique, the presence of irrelevant or redundant features can lead the system to focusing attention on the idiosyncrasies of the individual samples while losing sight of the broad relational picture that is essential for generalization beyond the training set. This problem is compounded when the number of observations is also relatively small, as in our case study. If the number of variables is comparable to the number of training patterns, the parameters of the model may become unstable and are unlikely to be replicated if the study were to be repeated. Feature selection seeks to remedy this situation by identifying a small subset of relevant features and using only them to construct the actual model. In this work, the selection of the most relevant features to be used in the classification phase is based on box plots. Box plots provide an excellent visual summary of many important aspects of a distribution and are useful for identifying its outliers (Massart et al., 2005).

The conceptual basis for using box plots in distinguishing the most relevant features for classification is that things can be distinguished from each other based on their inconsistency (Hsiao et al, 2009). Outliers can in fact be used as a primary method for pattern classification: the more outliers a parameter distribution has, the more that parameter will be useful in defining clusters in the feature space defined by the considered parameter while avoiding cluster overlapping. There are several steps in constructing a box plot. The first relies on the evaluation of the 25th, 50th, and 75th percentiles in the distribution of the 7 patterns.

Figures 4-7 show how these three statistics are used in our case study: for each extracted feature, we draw a box extending from the 25th percentile to the 75th percentile. The 50th percentile is drawn inside the box. We also put “whiskers” above and below each box to give additional information about the spread of data. Whiskers are vertical lines that end in a horizontal. They are drawn from the lowest and upper hinges to the lowest datum still within 1.5 Inter Quartile Range (IQR) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile, respectively (Massart et al., 2005). Finally, we represent outliers in our box plots by adding additional crosses beyond the whiskers.

From the analysis of Figures 4–7, the relevant features for each sensor were:

- S1: skewness and kurtosis
- S2: skewness
- S3: mean, standard deviation, kurtosis, clearance indicator, shape indicator and impulse indicator
- S4: skewness

Most of the four signal box plots highlight skewness and kurtosis spread distributions. According to this consideration and to keep controlled the computational burden of the approach, only these two fea-

tures were considered to be key features on which the classification of the degradation patterns would be based.

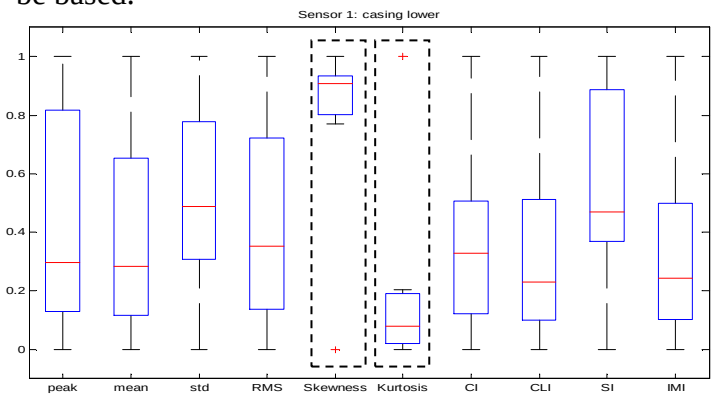

Figure 4. Box plots of the distributions of the 10 extracted features from sensor S1.

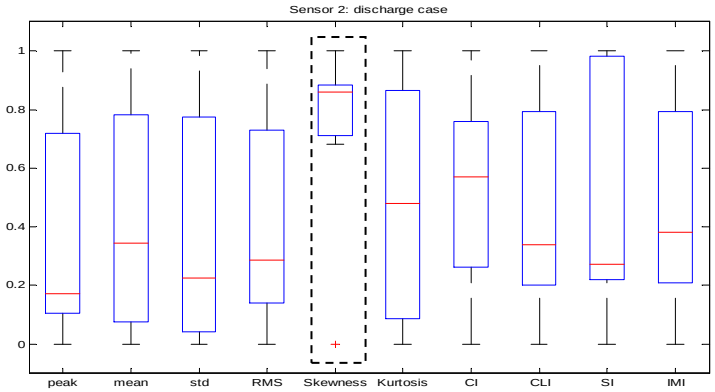

Figure 5. Box plots of the distributions of the 10 extracted features from sensor S2.

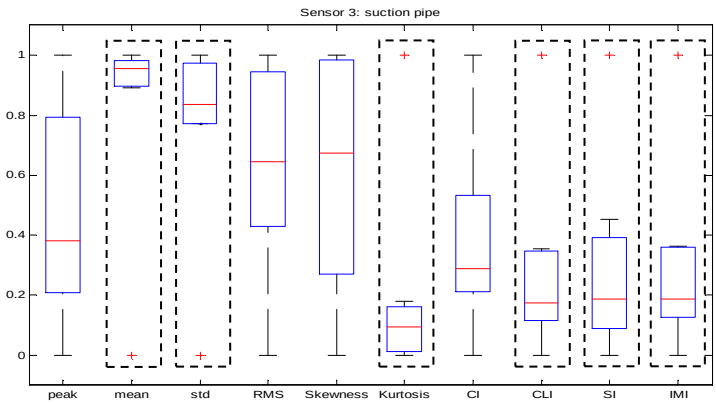

Figure 6. Box plots of the distributions of the 10 extracted features from sensor S3.

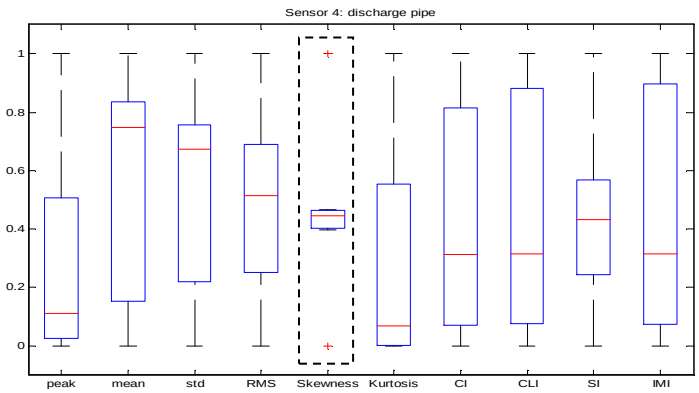

Figure 7. Box plots of the distributions of the 10 extracted features from sensor S4.

5 CLASSIFICATION

Fault detection may pose difficulties, because it entails the implementation of a classifier for labeling the component status as healthy or failed. In our application, the shortage of data forces us to resort to an unsupervised classification scheme for all the classifiers that in an attempt to improve the detection-classification performance are combined to estimate the status of the pumps (Freund et al., 1996; Schapire, 1999; Friedman, 2000). Figure 8 illustrates the basic framework for the ensemble scheme adopted. The key step was the formation of an ensemble of diverse classifiers from a single data set. In this work, four different classifiers were fed with different inputs taken from different sensors (S1, S2, S3 and S4), but all relative to the same degradation pattern. The single classifier results were then combined by two different methodologies (Friedman et al., 2000):

- Majority voting
  - Each ensemble member votes for one of the classes (the one with the largest membership value, see Section 5.1).
  - Predicts the class with the highest number of vote.
  - In case of equal number of votes, the class is labeled as uncertain.

- Weighted voting
  - Make a weighted sum of the votes of the ensemble members (Weights depend on the performance of each independent classifier).

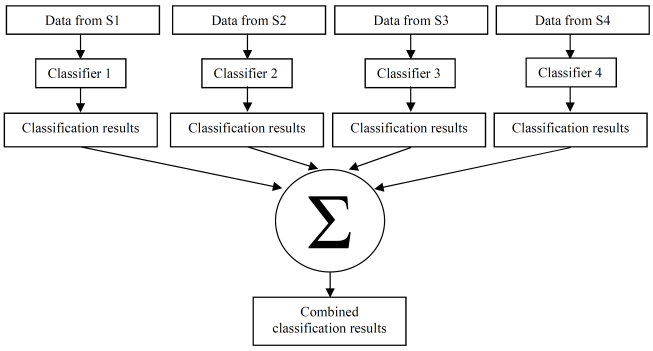

Figure 8. Scheme of combination of different classifiers.

5.1 *The Unsupervised Fuzzy C-Means algorithm*

Fuzzy C-Means (FCM) is one of the most popular fuzzy clustering methods (Bezdek, 1981; Leguizamon et al., 1996; Alata et al., 2008). The FCM method originated from hard C-Means clustering, allowing data points to belong to two or more clusters (Klir et al., 1995). The clusters emerge from the minimization of the following objective function:

$$J(N,C)=\sum_{i=1}^{N}\sum_{j=1}^{C}m_{ij}^{\varphi}d_{ij}^{2}(x_i,c_j) \quad (1)$$

where $J(N,C)$ is the sum of the square errors of the distance of each individual data point x_i, i=1,2,…,N, to the center c_j, j=1,2,…,C, of the given cluster (class) j. The minimization is done with respect to the membership m_{ij}^{φ} and the centers c_j. More specifically, $d_{ij}^{2}(x_i,c_j)$ is the square of the distance between x_i and c_j, whereas m_{ij}^{φ} is the degree of membership of x_i to cluster j. The value φ is any real number greater than 1, and it modulates the fuzziness of the clusters. Fuzzy partitioning is carried out through an iterative optimization of $J(N,C)$, with the update of memberships m_{ij}^{φ} and the cluster centers c_j by:

$$m_{ij}^{\varphi}=\frac{d_{ij}^{-2(\varphi-1)}}{\sum_{j=1}^{c}d_{ij}^{-2(\varphi-1)}} \quad (2)$$

$$c_j=\frac{\sum_{i=1}^{n}m_{ij}^{\varphi}x_i}{\sum_{i=1}^{n}m_{ij}^{\varphi}} \quad (3)$$

For further details, the interested reader may refer to (Bezdek, 1981).

6 RESULTS

The extracted features to be fed to the FCM classification algorithm were selected by box plot analysis in Section IV, where we justified the choice of skewness and kurtosis as important features. The classification phase identified two clusters that can be useful for labeling the degradation patterns as relative either to failed or to healthy pumps. In this case, the classification results are shown in Figure 9-12. By analyzing the skewness and kurtosis values of the considered degradation patterns plotted on the scatter plot of Figures 9-12, it turns out that the main differences between the two identified clusters (represented by circles and crosses) are:

- Circles have skewness values close to zero and lower kurtosis values.
- Crosses have skewness values far from zero and higher kurtosis values.

Based on engineering-based considerations we have decided that:

- Degradation patterns with skewness values close to zero, i.e., vibrational data normally distributed, are relative to healthy pumps (the flow of abrasive and erosive particles can only generate white noise on the measurements).
- Degradation patterns with skewness values far from zero, i.e., vibrational data not normally distributed, are working in anomalous conditions (failed components highly deform the parameter distributions).

Thus, circles are labels of degradation patterns for safe pumps, crosses indicate the class of failed pumps and stars are the clusters centers. The classification results based on the batch of four sensors, S1, S2, S3, and S4, from the 7 degradation patterns have been listed in Table III. Based on the FCM classifier algorithm, when using majority voting, the correctness of the classification was 86%, with an uncertain assignment percentage equal to 14%; whereas, using weighted voting, the correctness of the estimations reached 100%. In both cases, only one pattern is labeled as failed pump. However, the non-aggregated results of the classifiers shown in Figures 9-12 highlight that each classifier identifies a different number of patterns belonging to the two classes. In fact, in Figures 9 and 11, one pattern is labeled as failed, whereas in Figures 10 and 12, three and two patterns, respectively. This demonstrates that the aggregation of the four classifier results is less dependent on the specifics of a single classifier, showing that a combination of multiple classifiers may learn a more expressive concept class than a single classifier.

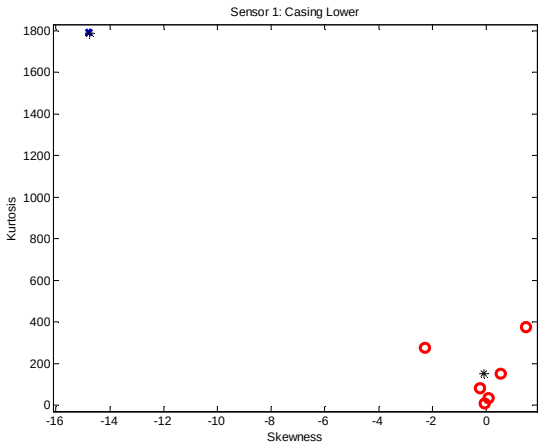

Figure 9. Classification results for sensor S1 (Classifier 1). Circles = Healthy, Crosses = Failed

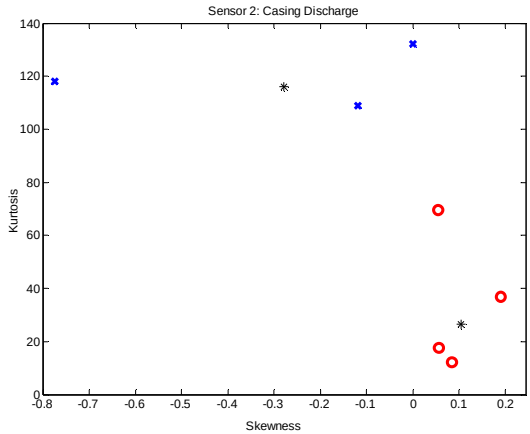

Figure 10. Classification results for sensor S2 (Classifier 2). Circles = Healthy, Crosses = Failed

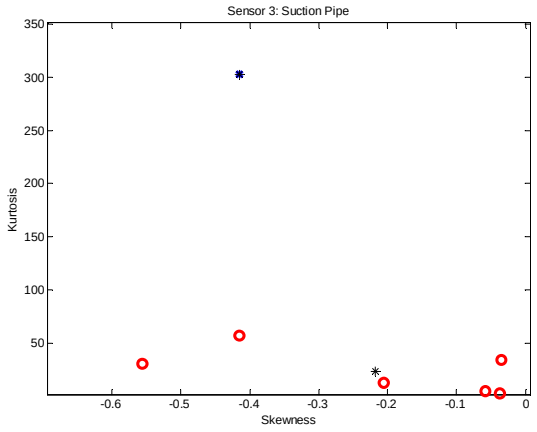

Figure 11. Classification results for sensor S3 (Classifier 3). Circles = Healthy, Crosses = Failed

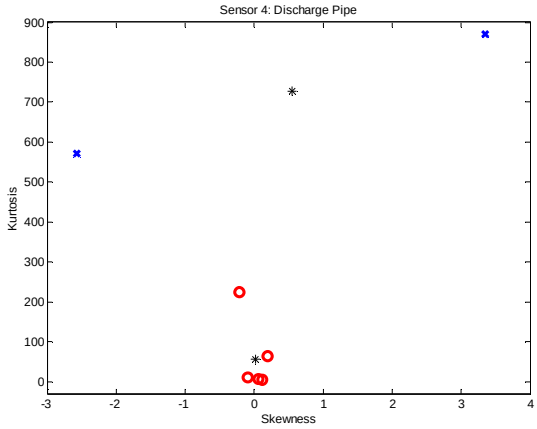

Figure 12. Classification results for sensor S4 (Classifier 4). Circles = Healthy, Crosses = Failed

Table 3. Classification Results Using Unsupervised FCM with Box Plots as Selecting Methods for the Input Features (H=healthy, F=failed, ?=uncertain).

Pattern	1	2	3	4	5	6	7	Correct	Uncertain
Majority voting	H	?	F	H	H	H	H	86%	14%
Weighted sum	H	H	F	H	H	H	H	100%	0%

7 CONCLUSIONS

In this work, we have presented a framework of analysis for assessing the wear status of pumps when available data is extremely limited. The method relies on an unsupervised clustering ensemble methods, based on FCM for classifying the available data. In particular, the adopted unsupervised FCM approach exploits the advantages of the automated generation of fuzzy rules, low computational burden, and the high-level, humanlike thinking and reasoning of fuzzy systems, which offer an appealingly powerful framework for tackling practical classification problems. Fault detection based on FCM allows building clusters with uncertain boundaries accommodating for different pump locations and different pump types and sizes. Moreover, the cluster centers identified by the FCM can turn out useful during online fault detection for classifying a new developing degradation pattern into healthy/failed clusters according to the distances of the feature values from the centers. The application of the framework (data collection, feature extraction, feature selection and classification) can be useful for industries to monitor the health of a machine prone to degradation and sporadic catastrophic breakdowns and dynamically plan equipment maintenance. However, further verification with additional real data is required for the framework to be of practical use in real industrial applications.

8 REFERENCES

Alata, M., Molhim, M., Ramini, A., 2008, Optimizing of Fuzzy C-Means Clustering Algorithm Using GA, *World Academy of Science, Engineering and Technology*, pp.224-229.

Bezdek, J.C., 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984, *Classification and regression trees*, Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Freund, Y., Schapire, R.E., 1996, Experiments with a New Boosting Algorithm, *Machine Learning: Proceedings of the Thirteenth International Conference*, pp. 148-156.

Friedman, J., 2000, Predictive Learning through Gradient Boosting, Keynote Address, *Seventeenth International Conference on Machine Learning*, Stanford University.

Friedman, J., Hastie, T., Tibshirani, R., 2000,"Additive Logistic Regression: a Statistical View of Boosting, *Ann. Statist.*, Volume 28, Number 2, 337-407.

Frith, R.H., Scott, W., 1996, Comparison of an External Gear Pump Wear Model with Test Data, *Wear*, Volume 196, Issues 1-2, pp. 64-71.

Guyon, I., Elisseeff, A., 2003, An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, pp.1157-1182.

Hancock, K.M., Zhang, Q., 2006, A Hybrid Approach to Hydraulic Vane Pump Condition Monitoring and Fault Detection, *Transactions of the American Society of Agricultural and Biological Engineers*, Vol. 49(4), pp. 1203-1211.

Hsiao, C., Chen, H., 2009, On Classification from Outlier View, http://arxiv.org/ftp/arxiv/papers/0907/0907.5155.pdf 2009.

Klir, G.J., Yaun, B., 1995, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall.

LaBour Pump Company, 1995, *LaBour Taber: Pump Wear*, http://www.peerlesspump.com/Pump Wear.pdf.

Leguizamon, S., Pelgrum, H., Azzali, S., 1996, Unsupervised Fuzzy C-Means Classification for the Determination of Dynamically Homogeneous Areas, *Proceedings of the VIII symposium of remote sensoring, Salvador do Brasil, INPE*, pp.851-856.

Lei, Y, He, Z., Zi, Y, 2009, Application of an intelligent classification method to mechanical fault diagnosis, *Expert Systems with Applications*, 36, pp. 9941–9948

Loh, W.Y., Shih, Y.S., 1997, Split selection methods for classification trees, *Statistica Sinica*, 7, pp. 815-840.

Marseguerra, M., Zio, E., Avogadri, P., 2004, Model Identification by Neuro-fuzzy Techniques: Predicting the Water Level in a Steam Generator of a PWR, *Progress in Nuclear Energy*, Volume 44, Issue 3, 2004, pp. 237-252

Massart, D.L., Smeyers-Verbeke, J., Capron, X., Schlesier, K., 2005, Visual Presentation of Data by Means of Box Plots, *Practical Data Handling, LC•GC Europe*, 18(4), pp. 215–218.

Mitchell, J.S., 1999, The History of Condition Monitoring and Condition-based Maintenance, *Sound and Vibration*, 33(11), pp. 21-28.

Ripley, B.D., 1996, *Pattern Recognition and Neural Networks*, Cambridge University Press.

Rumelhart, D.E., McClelland, J.L., 1986, *Parallel Distributed Processing*, Vol. 1. MIT Press, Cambridge, MA.

Shahrtash, S.M., Jamehbozorg, A., 2008, A Decision Tree–based Method for Fault Classification in Transmission Lines, *IEEE*.

Schapire, R.E., 1999, Theoretical Views of Boosting and Applications, in *Tenth International Conference on Algorithmic Learning Theory*.

Sheng, Y., Rovnyak, S.M., 2004, Decision Tree-Based Methodology for High Impedance Fault Detection, *IEEE Transactions on Power Delivery*, Vol. 19, No. 2.

Tian, J., Gao, M., Li, K., Zhou, H., 2007, Fault Detection of Oil Pump Based on Classify Support Vector Machine, *2007 IEEE International Conference on Control and Automation WeD5-4*, Guangzhou, China, May 30–June 1, 2007.

Tse, P.W., 2002, Maintenance Practices in Hong Kong and the Use of the Intelligent Scheduler, *Journal of Quality in Maintenance Engineering*, Vol. 8 Iss: 4, 2002, pp. 369–380.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., 2003, A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-based Methods, *Computers & Chemical Engineering*, Vol. 27, pp. 293-311.

Wang, J.P., Hu, H.T., 2006, Vibration-based Fault Diagnosis of Pump Using Fuzzy Technique, *Measurement*, Vol. 39, pp. 176-185.

Wang, H.Q., Chen, P., Fault Diagnosis of Centrifugal Pump Using Symptom Parameters in Frequency Domain, *Agricultural Engineering International: the CICR Ejournal*. Vol. 9, pp. 1-14.

Zadeh, L.A:, 1965, Fuzzy sets, *Information and Control*, 8, 1965, pp. 338-53.

Zhang, G.P., 2000, Neural Networks for Classification: A Survey, *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol. 30 (4).

Zio, E., 2007, Soft Computing Methods Applied to Condition Monitoring and Fault Diagnosis for Maintenance, *Reliability: Theory and Applications*, Vol. 3.

Zio, E., Gola, G., 2006, Neuro-fuzzy Pattern Classification for Fault Diagnosis in Nuclear Components, *Annals of Nuclear Energy*, Vol. 33, Issue 5, pp. 415-426.